AD-A153 379 EVIDENTIAL REASONING IN EXPERT SYSTEMS FOR IMAGE ANALYSIS(U) SYSTEM PLANNING CORP ARLINGTON VA T R THOMPSON FEB 85 ETL-0381 DACA72-84-C-0006 1/2

UNCLASSIFIED F/G 9/2 NL

1.0
1.1
1.25 1.4 1.6
4.5 5.0 5.6 6.3
2.8 3.2 3.6 4.0
2.5 2.2 2.0 1.8

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD-A153 379

ETL-0381

(1)

Evidential reasoning in expert systems for image analysis

Terence R. Thompson

System Planning Corporation
1500 Wilson Boulevard
Arlington, Virginia 22209

February 1985

Prepared for

U.S. ARMY CORPS OF ENGINEERS
ENGINEER TOPOGRAPHIC LABORATORIES
FORT BELVOIR, VIRGINIA 22060-5546

E T L

U.S. ARMY ENGINEER TOPOGRAPHIC LABORATORIES

ETL

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: ETL-0381	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): EVIDENTIAL REASONING IN EXPERT SYSTEMS FOR IMAGE ANALYSIS	5. TYPE OF REPORT & PERIOD COVERED: Final technical report June 1984 - December 1984
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Terence R. Thompson	8. CONTRACT OR GRANT NUMBER(s): DACA72-84-C-0006
9. PERFORMING ORGANIZATION NAME AND ADDRESS: System Planning Corporation, 1500 Wilson Boulevard, Arlington, Virginia 22209	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: U.S. Army Engineer Topographic Laboratories, Fort Belvoir, Virginia 22060-5546	12. REPORT DATE: February 1985
	13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Image Analysis, Expert Systems, Evidential Reasoning

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report documents efforts to understand approaches to evidential reasoning of use in application of expert-system or knowledge-based-system techniques to image analysis (IA). There is growing evidence that these techniques offer significant improvements in image analysis, particularly in the coordinated application of specialized algorithms.

This effort has four principal goals: (1) to clarify the basic issues in evidential reasoning (ER), (2) to provide a common framework for analysis,

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(3) to structure the ER process for major expert-system tasks in image analysis, and (4) to identify promising directions for further research.

This research was carried out in three major segments. The first segment structured the evidential-reasoning problem in a formal paradigm robust enough to be of practical use in design and construction of expert systems. It then formulated six important theoretical approaches in a parallel fashion in order to identify key assumptions, similarities, and differences.

The second segment applied each of the ER approaches to three important tasks for expert systems in the domain of image analysis. This segment concluded with an assessment of the strengths and weaknesses of each approach.

The third segment addressed promising directions for further research. It reviewed current results and identified important questions bearing on successful application of expert-system technology to image analysis.

PREFACE

This document was prepared under contract DACA72-84-C-0006 for the U.S. Army Engineer Topographic Laboratories, Fort Belvoir, Virginia, by Systems Planning Corporation, Arlington, Virginia. The Contracting Officer's Representative was Mr. Robert S. Rand.

A-1

CONTENTS

I: EXECUTIVE SUMMARY

A. PURPOSE AND SCOPE

This report documents efforts to understand approaches to evidential reasoning that may be useful in application of expert-system techniques to image analysis (IA). These techniques offer significant improvements in image analysis, particularly in the coordinated application of specialized algorithms.

A central element in such expert-system (ES) applications is the handling of evidence. In most tasks, evidence accumulates over time to dynamically affect uncertainties, so that the decision preferred earlier may differ from the one preferred later. However, delaying a decision is often not feasible, since this may foreclose opportunities or increase costs. Thus, it is important to understand how accumulating evidence will affect the decision process in the face of uncertainty.

There is no general consensus on how best to attack evidential-reasoning (ER) problems, particularly in expert-system applications. Several approaches have evolved, but they have their roots in diverse fields, such as statistics and philosophy, and have neither a common terminology nor a common set of assumptions.

The effort documented here has four principal goals: (1) to clarify the basic issues in evidential reasoning, (2) to provide a common framework for analysis, (3) to structure the ER process for major expert-system tasks in image analysis, and (4) to identify directions for further research. Its scope has been bounded in the following ways:

- Effort has been spread evenly across the spectrum of ER approaches; each has been treated to approximately the same level of detail in order to provide a uniform view of relative applicability.

- Only low-level vision through the level of the primal sketch (M3) is considered, although extensions to high-level vision are certainly possible.
- Only rule-oriented techniques for expert-systems are considered; extensions to data-oriented, object-oriented, and procedure-oriented techniques are possible.

The approach used to attain these goals within the bounds cited is summarized below.

B. APPROACH

This research was carried out in three major segments. The first is primarily concerned with theories of evidential-reasoning, the second with applications in image-analysis expert systems, and the third with current results and further research.

The first segment of the research takes place in two steps. In the first step, it structures the evidential-reasoning problem in a formal paradigm robust enough to be of practical use in design and construction of expert systems. The elements of the paradigm are:

- Background Elements
- Observation Reports
- Updating Mechanism
- Decision Mechanism.

In the second step, this segment formulates six important theoretical approaches in a parallel fashion in order to identify key assumptions, similarities, and differences. The six approaches are:

- Classical Bayes
- Convex Bayes
- Dempster-Shafer
- Kyburg
- Neyman-Pearson
- Possibility.

This segment results in parallel formulations of the ER approaches and a discussion of points of correspondence and incommensurability.

The second segment of the research applies the ER approaches to three important tasks for expert systems in the domain of image analysis. The tasks discussed are:

- Diagnosis - the inference of system behavior from various reports
- Integration - the meaningful combination of a number of disparate inputs into a smaller number of outputs
- Control - the choice of actions that influence system behavior.

Sample tasks are constructed for each type, and the application of each ER approach to each sample task is discussed. This segment concludes with an assessment of the strengths and weaknesses of each approach.

The third segment of the research addresses directions for further effort. It first summarizes the results of the current effort and then identifies important questions that bear on successful application of expert-system technology to image analysis.

C. RESULTS

The evidential-reasoning research reported in Chapter II can be summarized as follows:

- The evidential-reasoning problem can be formulated in terms of a four-part paradigm:
    - -- Background Elements - This portion of the paradigm contains a definition of the domain of discourse, that is, of the world-model to which we shall apply the ER process. It also contains current knowledge of that world-model to which we shall apply the ER process. It also contains current knowledge of that world including, possibly, knowledge of the cost of various actions in that world. Knowledge is described in terms of belief states.
    - -- Observation Reports - This portion of the paradigm describes the structure and content of reports about the external world that are the raw material for revision of the knowledge embedded in the background.
    - -- Updating Mechanism - This portion of the paradigm describes the assumptions, rules, and algorithms used to revised knowledge upon receipt of observation reports.

-- Decision Mechanism - This portion of the paradigm describes the assumptions, rules, and algorithms used to choose among various courses of action given revised knowledge of the world.

- Each of the six major approaches can be expressed in terms of the four-part paradigm.
- Major similarities in the ER approaches are found in two background elements:

-- Structure of the algebra of statements (but not necessarily the content)

-- The loss function.

- Major differences in the ER approaches are found in several components:

-- Structures given to belief states (points, intervals, convex sets, fuzzy sets)

-- Updating algorithms (Bayes' Theorem, Dempster's Rule, principles of direct inference, confidence intervals, fuzzy combination)

-- Decision algorithms (expected loss on point-valued p-functions, expected loss on intervals or convex sets, fuzzy decision rules).

The research into application of evidential-reasoning approaches to expert-system tasks in image analysis reported in Chapter III can be summarized as follows:

- Major expert-system tasks in this domain are: (1) diagnosis, the inference of system behavior from data on system processes, (2) integration, the meaningful combination of a number of disparate inputs into a smaller number of outputs, and (3) control, the choice of actions that influence system behavior.
- Each of the ER approaches can be applied to sample tasks from these three categories. Several strengths and weaknesses can be identified:

-- Interval and convex-set representations of belief states may be useful in complex ES tasks (e.g., control), but do so at the expense of added complexity.

-- Specialized decision procedures must be developed in order to make practical use of these robust representations.

-- Criteria of evidential relevance are being developed, but require practical application for assessment.

-- Imprecise linguistic terms may be characterized by fuzzy sets, but this also requires practical application for assessment.

Directions for further research include efforts to: (1) develop a prototype image-analysis expert system for application to current concerns of the Engineer Topographic Laboratories (ETL), (2) compare several ER techniques in direct application to one or more detailed image-analysis tasks, (3) develop rules for specific image-analysis tasks, (4) investigate the relative utility of various types of rule-based control systems, and (5) investigate the utility of trainable or learning expert-systems for image analysis.

II: PARALLEL FORMULATION OF RELEVANT EVIDENTIAL-REASONING THEORIES

A. INTRODUCTION

This portion of our research serves two purposes. First, it structures the evidential-reasoning problem in a paradigm robust enough to be of practical use in design and construction of expert systems. Second, it formulates six important theoretical approaches in a parallel fashion in order to identify key assumptions, similarities, and differences.

Effort applied to this part of our research has been spread evenly across the spectrum of ER approaches. Each approach has been treated to approximately the same level of detail in order to provide a uniform view of relative applicability.

Section B structures the ER problem. Sections C and D formulate and compare the six theoretical approaches. Section E summarizes results of this chapter.

B. THE EVIDENTIAL-REASONING PROBLEM

1. General Description

The problem of evidential reasoning is a very general one, and may be formulated as follows:

- Given reports about the world, and a set of current beliefs about the world, how shall I revise my beliefs as new reports are received?

Reports may range from the simple to the complex in referring to various objects or sets of objects in the world. They also may refer to events and may contain various uncertainties. Reports may even refer elliptically to ill-defined sets. Beliefs also range from the simple to the complex, and have a notoriously obscure structure.

Of course, since we seek to construct expert systems to aid in certain relatively well-defined image-analysis tasks, all of the complications implicit in the question above need not be explored here. However, a central element in such expert-system applications as diagnosis, integration, and control is the handling of evidence. In such tasks, evidence accumulates over time to dynamically affect uncertainties, so that the decision preferred earlier may differ from the one preferred later.

There is no general consensus on how best to attack evidential-reasoning problems, particularly in expert-system applications. Several different theoretical approaches have evolved, but they have their roots in diverse fields, such as statistics and philosophy, and have neither a common terminology nor a common set of assumptions. This makes it difficult to answer such questions as:

- What are the rules for structuring the reports about the world that feed raw material into the evidential-updating schemes advocated by each theoretical approach?
- What are the constraints on ER that are implicit (and explicit) in application of each of the approaches? More broadly, what models of ER are implicitly and explicitly advocated by each approach?

Such concerns lead us to seek a structured paradigm broad enough to encompass the models associated with each approach. This paradigm will be used to identify and compare assumptions, rules, and constraints.

2. A Structured Paradigm

The structured paradigm for the ER process that we shall use throughout the remainder of this report has four components:

- Background Elements - This portion of the paradigm contains a definition of the domain of discourse, that is, of the world-model to which we shall apply the ER process. It also contains current knowledge of that world including, possibly, knowledge of the cost of various actions in that world. Knowledge is described in terms of belief states.

- Observation Reports - This portion of the paradigm describes the structure and content of reports about the external world that are the raw material for revision of the knowledge embedded in the background.
- Updating Mechanism - This portion of the paradigm describes the assumptions, rules, and algorithms used to revise knowledge upon receipt of observation reports.
- Decision Mechanism - This portion of the paradigm describes the assumptions, rules, and algorithms used to choose among various courses of action given revised knowledge of the world.

Figure II-1 shows the inter-relationships of components of the paradigm.

Some of the research literature excludes decision-making from the ER process. It is included here for two reasons: (1) the image-analysis tasks envisioned for expert systems will generally involve decisions of one sort or another, and (2) the structure of the other components is intimately linked, in most cases, with the decision process.

The following section discusses each ER approach within the common framework provided by the structured paradigm. Section D compares them.

C. THEORETICAL APPROACHES

We will discuss six major approaches to evidential reasoning:

- Classical Bayes - based upon point or interval represenations of belief states and Bayes' Theorem
- Convex Bayes - based upon convex sets and Bayes' Theorem
- Dempster-Shafer - based upon mass functions and Dempster's Rule of Combination
- Kyburg - based upon interval representations and direct inference
- Neyman-Pearson - based upon confidence intervals
- Possibility - based upon fuzzy sets and degree-of-membership functions.

Each will be presented separately in terms of the structured ER paradigm described in Section B. A comparative analysis will be carried out in Section D.

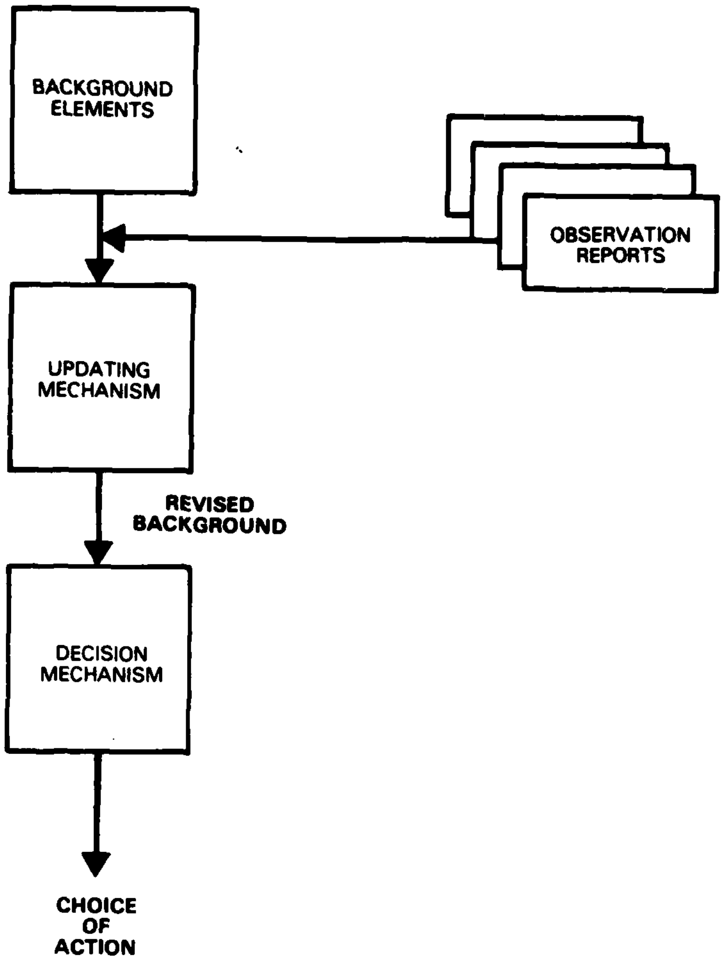


FIGURE II-1.
STRUCTURED PARADIGM FOR EVIDENTIAL REASONING

1. Approach 1: Classical Bayes

a. Background Elements

The background in this approach consists of three elements: (1) an algebra of statements, (2) a probability function defined over this algebra, and (3) a utility function defined over the same algebra. The algebra defines the domain of discourse, the probability function assigns degrees of belief to elements of the domain, and the utility function provides a means of reaching decisions in the domain when coupled with the decision mechanism.

The algebra used in the classical Bayes approach is known as a Lindenbaum-Tarski (LT) algebra. It consists of base elements, operators, and propositions entailed by application of the operators to the base elements.

The base elements are variously known as atoms, states of affairs, or possible worlds. They are assumed to be mutually exclusive, so that application of the disjunctive operator alone expands the base elements into the set of all possible legal statements about the domain of discourse. In our discussion, we shall refer to the mutually exclusive elements as base elements and to the legal statements as atoms.

For example, if there are four mutually exclusive base elements labelled "1", "2", "3", and "4", then the set of legal statements has the following members:

(null)

(1) (2) (3) (4)

(1 v 2) (1 v 3) (1 v 4) (2 v 3) (2 v 4) (3 v 4)

(1 v 2 v 3) (1 v 2 v 4) (1 v 3 v 4) (2 v 3 v 4)

(1 v 2 v 3 v 4)

In general, there will be N_n legal statements when there are n base elements, where

$$N_n = \sum_{p=0}^{n} \binom{n}{p}$$

$$= \sum_{p=0}^{n} \frac{n!}{(n-p)!\ p!}$$

$$= 2^n$$

Thus there will be 16 statements if there are 4 base elements, 256 statements given 8 base elements, and so on.

The second major element of the background is a probability function defined over the algebra of statements and obeying the following axioms:

$$0 < p(x) < 1$$

$$p(x \vee y) = p(x) + p(y), \text{ if } x \text{ and } y \text{ are mutually exclusive.}$$

The sum of the probabilities assigned to the base elements is required to be 1.

The probability function assigns numbers to the legal statements based upon these axioms. For example, if the probabilities assigned to the four base elements are each 0.25, then the legal statements have the following p-values:

(null)
0.0

(1) (2) (3) (4)
.25 .25 .25 .25

(1 v 2) (1 v 3) (1 v 4) (2 v 3) (2 v 4) (3 v 4)
.50 .50 .50 .50 .50 .50

(1 v 2 v 3) (1 v 2 v 4) (1 v 3 v 4) (2 v 3 v 4)
.75 .75 .75 .75

(1 v 2 v 3 v 4)
1.0

The updating mechanism discussed below controls the manner in which these p-values change as evidence is received.

The third major element of the background is a utility function defined over the algebra of statements. This function is ordinarily construed as a loss function; it gives the loss, l_{ij}, incurred when the i^{th} action is taken in the face of the state of nature corresponding to the j^{th} base element in the algebra (B5, C2, J1).

For example, if there are three possible actions and four base elements, lij could be represented by the following matrix of i rows and j columns:

$$l_{ij} = \begin{vmatrix} 4 & 3 & 2 & 1 \\ 2 & 0 & 1 & 9 \\ 3 & 6 & 9 & 2 \end{vmatrix} \quad .$$

The decision mechanism discussed below controls the manner in which the loss function is used to indicate which action should be taken.

b. <u>Observation Reports</u>

The observation reports are direct assignments of new p-values to elements of the algebra of statements. That is, they assign a number or numbers to certain propositions that may be construed as a new degree of belief in the truth-value of that proposition. The assignment of this new p-value causes a re-assignment of p-values to other statements in the algebra via the updating mechanism discussed in the following section.

There are several ways in which this direct assignment of new p-values may be viewed:

- Each observation report consists of the assignment of a single p-value of 1.0 to some element in the algebra of statements.
- Each observation report consists of the assignment of a single p-value in the interval [0,1] to some element in the algebra.
- Each observation report consists of the assignment of two p-values in the interval [0,1] to some element in the algebra. These are construed as lower and upper p-values for the element.

The primary effect of these different views is upon the size of the algebra of statements. The number of statements required is largest under the first view, since we must have a single statement corresponding to each and every possible observation (parameter value at a certain pixel, value read on a meter, etc.). The other views allow us to use smaller numbers of statements, since we may map several observations onto a single statement. Ordinarily, only the first view is utilized in the classical Bayes approach.

We shall consider a simple example in order to demonstrate the effect upon the size of the algebra of statements. Consider a situation in which the task is to differentiate between a river and a road based upon measurements of the brightness of candidate topographic features. Suppose further that the brightness measurement is simply one of three possibilities: low, moderate, or high.

The propositions used to form the algebra of statements would be:

S1 - "The feature is a road."
S2 - "The feature is a river."
S3 - "The brightness of the feature is low."
S4 - "The brightness of the feature is moderate."
S5 - "The brightness of the feature is high."

Under each of the views of observation reports, one simply assigns p-value(s) to one of these statements and then uses the updating mechanism to modify the other p-values.

Under the first view, the algebra must include statements that correspond to every possible observation, since the nature of an observation is to assign some particular statement a p-value of 1.0. In terms of the present example, the addition of propositions like the following would be required if we are actually receiving observations consisting of cloud-obscured reflectances:

S6 - "The cloud-obscured reflectance measurement is 1.0."
S7 - "The cloud-obscured reflectance measurement is 2.0."

S8 - "The cloud-obscured reflectance measurement is 3.0."

S9 - "The cloud-obscured reflectance measurement is 4.0."

S10- "The cloud-obscured reflectance measurement is 5.0."

This view, in essence, leads one to embed uncertainties in the matrix of conditional probabilities. That is, the web of inference leads to p-values less than 1.0 for some statements based upon unitary p-values for other statements. We shall call this sort of uncertainty "inferential uncertainty."

Under the second view, one could use the shorter set of statements (S1 through S5), assign a p-value between 0.0 and 1.0 to one of them, and use the updating mechanism to modify the others. This means that the web of inference is prepared to operate not only with inferential uncertainty, but also with a second sort of uncertainty, one that we shall call "evidential uncertainty." Understanding the conditions for which this is advantageous is an area of ongoing research (T1).

The third view of observation reports, like the second, allows the use of both inferential and evidential uncertainty. It adds the feature of lower and upper p-values; these, in a sense, spread the evidential uncertainty. As before, situations in which this is advantageous are being explored in other research.

c. Updating Mechanism

The classical Bayesian approach rests upon Bayes' Rule for calculating posterior probabilities of states of nature from two items: (1) prior probabilities on those states, and (2) conditional probabilities for evidence given certain states of nature. In symbolic form,

$$P(S_i/E_j) = \frac{P(S_i)\ P(E_j/S_i)}{\sum_k P(S_k)\ P(E_j/S_k)} \qquad (1.1),$$

where

$P(S_i/E_j)$ = posterior probability of state S_i given evidence E_j

$P(S_i)$ = prior probability of state S_i (i.e., before evidence is taken into account)

$P(E_j/S_i)$ = conditional probability of evidence E_j given state S_i.

Given a probability or degree-of-belief distribution on the evidence, $P(E_k)$, we then compute the current p-value for each state of nature from the posterior probabilities and the evidential p-values, $P(E_k)$, according to

$$P_{cur}(S_i) = \sum_k P(E_k)\, P(S_i/E_k) \qquad (1.2),$$

where we assume that the distribution on the evidence is normalized to one. Variations on this approach are possible depending upon the structure of the algebra of statements. Note that the formula used here is compatible with both the first and second approaches to observation reports. Figure II-2 provides an overview of classical Bayesian updating.

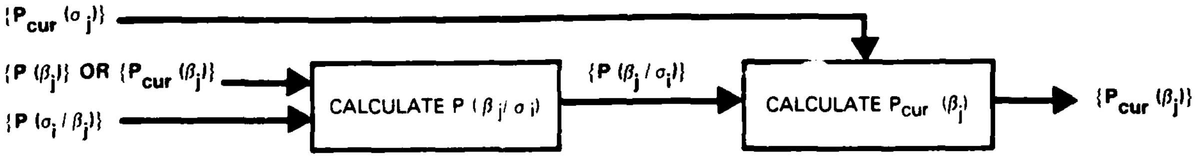


FIGURE II-2.
OVERVIEW OF CLASSICAL BAYESIAN UPDATING

509-2-14-85-5

In terms of the algebra of statements discussed in this section, there are no formal differences between any of the statements in the algebra. That is, using statements S1 through S5 of our current example, observation reports can be received for any one of the statements, even though the meaning we have attached to S3 through S5 would lead us to classify these three statements as evidence and the other two as states of nature.

Furthermore, we must remember that the algebra does not consist only of propositions S1 through S5. It is actually made up of the following mutually exclusive base elements:

BE1 = (S1 & S3)	BE4 = (S2 & S3)
BE2 = (S1 & S4)	BE5 = (S2 & S4)
BE3 = (S1 & S5)	BE6 = (S2 & S5) .

Since there are 6 base elements, there are 26 or 64 statements in the algebra.

Let us continue to flesh out the example. We need a set of conditional p-values:

$$PCON_{ij} = \begin{vmatrix} 1.0 & 0.0 & 0.9 & 0.5 & 0.3 \\ 0.0 & 1.0 & 0.1 & 0.5 & 0.9 \\ 0.9 & 0.1 & 1.0 & 0.0 & 0.0 \\ 0.5 & 0.5 & 0.0 & 1.0 & 0.0 \\ 0.3 & 0.9 & 0.0 & 0.0 & 1.0 \end{vmatrix} ,$$

where each element in this matrix of i rows and j columns gives the conditional p-value of the i^{th} statement in the algebra given that the p-value of the j^{th} statement is 1.0. The values in this matrix are assumed to be constant throughout the updating process described here.

We also need a matrix of the current p-values:

$$\mathrm{PCUR} = \begin{vmatrix} P_{cur}(S_i) \\ P_{cur}(S_2) \\ P_{cur}(S_3) \\ P_{cur}(S_4) \\ P_{cur}(S_5) \end{vmatrix},$$

where each element in this matrix of i rows gives the current p-value of the ith statement in the algebra. The values in this matrix are revised each time that a new observation report arrives. Each one is initially equal to the a priori value, $P(S_i)$.

Some writers divide the statements in the algebra into two distinct classes: hypotheses and evidence (D4, D5). Hypotheses are often called states of nature, while evidence is often termed measurements. In any case, the basic idea is that there is a directionality in the web of inference: we reason from evidence to hypotheses. We shall term this the hierarchical approach.

This approach reduces the dimensionality of updating calculations. In the context of our example, observation reports now can only be received on statements S3, S4, and S5. The conditional p-values for the evidence given the hypotheses are:

$$P(E_i/H_j) = \begin{vmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \\ 0.3 & 0.9 \end{vmatrix},$$

where each element in this matrix of i rows and j columns gives the conditional p-value of the i^{th} evidential statement in the algebra given that the p-value of the j^{th} hypothesis is 1.0. The values in this matrix are assumed to be constant throughout the updating process.

The matrix of the current p-values for hypotheses becomes:

$$PCUR(H_i) = \begin{vmatrix} P_{cur}(H_1) \\ P_{cur}(H_2) \end{vmatrix}$$

where each element in this matrix of i rows gives the current p-value of the i^{th} hypothesis in the algebra. The values in this matrix are revised each time that a new observation report arrives.

The hierarchical approach may be extended so that evidential statements can serve in one of two different roles in the web of inference. First, they may serve as evidence for the hypothesis set as discussed above. Second, they may serve as evidence for other evidential statements, rather than for the hypothesis set.

There may be an advantage in taking some form of hierarchical approach. First, the inferential relationships between statements in the algebra are made more explicit than they are in the undifferentiated algebra. Second, the computational burden associated with each updating cycle may be lessened in that the effects of an observation report are limited to portions of the hierarchy explicitly connected with the statement set that is the subject of the report.

For either the non-hierarchical or hierarchical approach, if we use the first interpretation of observation reports, the updating mechanism operates just once. We have an a priori set of $PCUR(H_i)$, we receive an observation report that assigns a p-value of 1.0 to one of the evidential statements, and we calculate an new set of $PCUR(H_i)$ using equations (1.1) and (1.2).

Under the second interpretation of observation reports, p-values may have values less than 1.0 and may therefore change over time. This would enable the updating cycle to occur more than once. For example, the p-value for evidential statement S3 might first be received as 0.5, then later as 0.7, and still later as 0.9.

The structure of legal evidential statements as embedded in the algebra of statements is an important issue. Certainly some evidential statements come in sets in the sense that they correspond to one observation that can have multiple outcomes, such as the reading on a digital meter. In that case, the observation report is a set of p-values, one for each evidential statement in the set. For example, the report on S3, S4, and S5 might be:

$$OR1 = \{0.5, 0.3, 0.2\} \quad .$$

This would lead to an updating cycle. Additional cycles might be induced upon receipt of reports like the following:

$$OR2 = \{0.7, 0.2, 0.1\}$$
$$OR3 = \{0.9, 0.1, 0.0\} \quad .$$

The updating cycles would continue as long as new reports were received. Other important issues are the effect of different sequences of reports upon the evolution of p-values and optimal control of these sequences. These lie beyond the scope of the present effort, but will be addressed in subsequent work.

d. Decision Mechanism

Given that the updating mechanism provides us with p-values for the states of nature, and given that the background contains the loss function, we can formulate the expected loss of the i^{th} action as follows:

$$EL_i = \sum_j (l_{ij} * PCUR_j) \qquad (1.3),$$

where the summation is over the j states of nature.

The general Bayesian decision function is simply to chose, whenever a decision is required, the action that gives the minimum value of EL_i. This is, of course, based upon the current set of $PCUR_j$ and the nature of the loss function.

The precise manner in which the loss function has been constructed may affect the decision, but discussion of this construction is beyond the scope of this effort. We simply mention here that the loss function matrix can be made up of absolute losses or relative losses. An absolute loss is the cost of taking the i^{th} action in the face of the j^{th} state of nature, while a relative loss for the same action-state pair is the difference between the absolute loss of that pair and the minimum of the absolute losses incurred by all actions in the face of the j^{th} state of nature. Relative losses are often termed "regrets."

Similar remarks can be made concerning the construction of the table of actions. The actions can range from very simple to very complex. A simple table of actions might be:

A1 - Take the pixel under consideration to be part of an edge feature, if the state of nature is A.

A2 - Take the pixel under consideration not to be part of an edge feature, if the state of nature is B.

A more complex table of actions might be dependent upon the number and structure of observation reports. For example, we might find actions like the following in such a table:

A3 - Take the pixel under consideration to be part of an edge feature, if observation report O_1 has been received.

A4 - Take the pixel under consideration not to be part of an edge feature, if reports O_2 or O_3 have been received.

Elaboration of the various complex tables of actions is highly dependent upon the specific task being addressed, as will be seen in Chapter III.

Decision theory offers an alternative to the approach based upon current p-values for states of nature. In the decision-theoretic approach, one constructs decision strategies based upon an assumption of ignorance concerning the state of nature. Discussion of the construction and comparison of such decision strategies (e.g., minimax loss and minimax regret) is also beyond the scope of the current effort.

We conclude this section with a brief summary of the classical Bayesian approach (Table II-1). This table may be compared with similar tables for the other approaches.

TABLE II-1
SUMMARY OF CLASSICAL BAYESIAN APPROACH

- Background Elements
  - -- Algebra of statements
  - -- Probability function defined over algebra of statements
  - -- Loss function defined over algebra of statements and embodying actions relevant to that algebra.
- Observation Reports
  - -- Association of p-value with statements in the algebra.
- Updating Mechanism
  - -- Via Bayes' Theorem, calculate posterior probabilities based on prior and conditional probabilities. In symbolic form,

$$P(S_i/E_j) = \frac{P(S_i)\ P(E_j/S_i)}{\sum_k P(S_k)\ P(E_j/S_k)} .$$

  - -- Use the posterior probabilities and the p-values for evidence to calculate current p-values on states via

$$P_{cur}(S_i) = \sum_k P(E_k)\ P(S_i/E_k).$$

- Decision Mechanism
  - -- Choose the action that minimizes the loss function using the current probability function.

2. Approach 2: Convex Bayes

a. Background Elements

The background in this approach, like the classical Bayes approach, consists of three elements: (1) an algebra of statements, (2) a probability function defined over this algebra, and (3) a utility or loss function defined over the same algebra. As before, the algebra defines the domain of discourse, the probability function assigns degrees of belief to elements of the domain, and the utility function provides a means of reaching decisions in the domain.

The convex Bayes approach also uses an LT algebra. It consists of base elements, operators, and propositions entailed by application of the operators to the base elements. As before, the base elements are assumed to be mutually exclusive, so the application of the disjunctive operator expands them into the set of all possible legal statements about the domain of discourse.

The probability function in the convex Bayes approach differs in a significant way from the function in the classical approach. Here the function is a convex set of p-functions (L2). That is, the belief state is not characterized by a single function but by a set of functions having the property of convexity: the set contains every linear combination of any two members of the set.

In general, if there are n base elements, the belief state will correspond to a domain in a space of (n-1) dimensions, since the n^{th} component of the belief state can be determined if (n-1) components are known. For example, suppose that there are three base elements. The belief state is then a domain in the the two-dimensional space depicted in Figure II-3. The region indicated depicts a possible belief state, $\{B_i\}$.

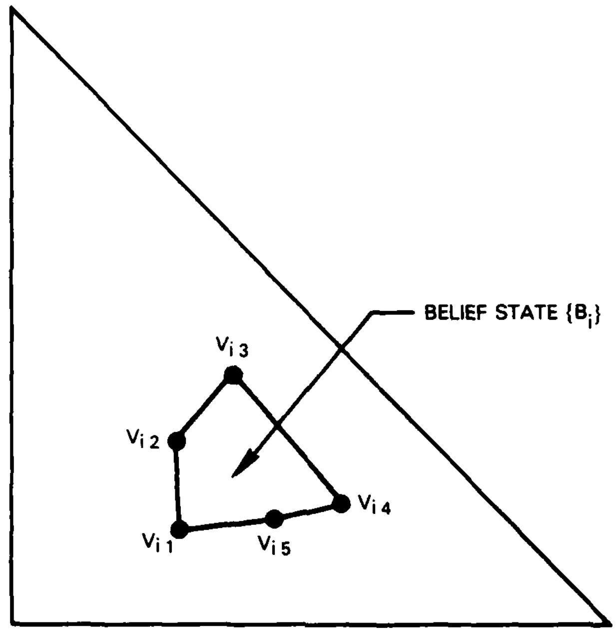


FIGURE II-3.
REPRESENTATION OF BELIEF STATE FOR THREE BASE ELEMENTS IN THE CONVEX BAYES APPROACH

509-2-14-85-8

It is a five-sided polygon with vertices v_{i1} through v_{i5}, so that it can be represented compactly by its vertices as follows:

$$V(B_i) = \begin{vmatrix} v_{i1} \\ v_{i2} \\ v_{i3} \\ v_{i4} \\ v_{i5} \end{vmatrix} = \begin{vmatrix} 0.4 & 0.2 & 0.4 \\ 0.4 & 0.3 & 0.3 \\ 0.5 & 0.4 & 0.1 \\ 0.6 & 0.2 & 0.2 \\ 0.7 & 0.2 & 0.1 \end{vmatrix} .$$

It must be remembered that this matrix represents only the vertices of B_i; any point on the boundaries or within them is also a member of $\{B_i\}$. In addition, the belief state may have any number of vertices, but must have at least one.

b. Observation Reports

The convex Bayes approach, like the classical approach, construes the observation reports as direct assignments of new p-values to elements of the algebra of statements. That is, the reports assign a number to certain propositions that may be construed as a new degree of belief in the truth-value of that proposition. The assignment of this new p-value causes a re-assignment of p-values to all other statements in the algebra via the updating mechanism discussed in the following section.

There are now four ways, rather than the three discussed with reference to the classical approach, in which this direct assignment of new p-values may be viewed:

- Each observation report consists of the assignment of a single p-value of 1.0 to some element in the algebra of statements.
- Each observation report consists of the assignment of a single p-value in the interval [0,1] to some element in the algebra.
- Each observation report consists of the assignment of two p-values in the interval [0,1] to some element in the algebra. These are construed as lower and upper p-values for the element.
- Some observation reports consist of the assignment of two or more linked bounds on the convex set of p-values. These bounds are linked in the sense that they jointly specify limits on the set.

As before, the primary effect of these different views is upon the size of the algebra of statements. The updating mechanism remains similar, as we shall see below. However, the decision mechanism based upon either single or multiple lower and upper p-values remains problematic.

c. Updating Mechanism

In essence the updating mechanism in the convex Bayes approach operates like the updating mechanism of the classical Bayes approach. The key difference is that the entire convex set of functions comprising the belief state is used, rather than a single function.

As before, we use Bayes' Theorem to calculate posterior probabilities based on prior and conditional probabilities. In symbolic form,

$$P_r(S_i/E_j) = \frac{P_r(S_i)\, P_r(E_j/S_i)}{\sum_k P_r(S_k)\, P_r(E_j/S_k)} \qquad (2.1),$$

where P_r is the r^{th} member of the convex set of probability functions.

This formula is used as follows:

- Assume evidence E_j has been presented.
- For each value of r, find the set of $P_r(S_i/E_j)$, using the set of $P_r(S_k)$, the set of $P_r(E_j/S_k)$, and the formula above. The values of k may be written as $\{1,2,\ldots,i,\ldots,k_{max}\}$. These sets each have k_{max} members.
- The output is the convex set of posterior current probability functions, that is, the convex set of the sets of $P_r(S_i/E_j)$.
- When new evidence E_n is presented, repeat this procedure using E_n in place of E_j, and the set of $P_r(S_i/E_j)$ in place of the set of $P_r(S_k)$ (L2, pp. 83-84).

Each new evidential input thus induces a mapping from one convex set of p-functions to another convex set.

It is clear that the computational burden of the updating mechanism is increased by use of the convex set of p-functions in place of a single p-function. Little work has been done in actual computation of updated convex belief states, so the extent of this burden is unclear at this time.

d. Decision Mechanism

Upper and lower probabilities for some statement in the algebra can be taken from the convex set of $P_r(S_i/E_j)$ using the technique of supporting lines, planes, or hyper-planes (L2, pp.196-198). However, no general procedure exists to handle upper and lower bounds in a utility function.

One method of attack is to suppose that the decision indicated is the one that minimizes the expected loss as was done in the classical approach. In this case we seek to minimize

$$EL_i = \sum_j (l_{ij} * PCUR_j) \quad (2.2),$$

where the summation is over the j states of nature. Using the convex set of $P_r(S_i/E_j)$, we derive upper and lower bounds on each $PCUR_j$ so that, for each action, there are now upper and lower bounds on the expected loss. If $PCUR_j$ is bounded by $PCURU_j$ and $PCURL_j$, then we might say that EL_i lies between $ELMIN_i$ and $ELMAX_i$, where

$$ELMIN_i = \sum_j (l_{ij} * PCURL_j) \quad (2.3),$$

$$ELMAX_i = \sum_j (l_{ij} * PCURU_j) \quad (2.4).$$

Such intervals for different actions will, in general, overlap. No generally accepted method for choice of actions has yet been developed, although minimax techniques have been explored by Kyburg and Levi.

We conclude this section with a brief summary of the convex Bayesian approach (Table II-2). This table may be compared with similar tables for the other approaches.

TABLE II-2
SUMMARY OF CONVEX BAYESIAN APPROACH

- Background Elements
  - -- Algebra of statements
  - -- Convex set of probability functions defined over algebra of statements
  - -- Utility function (not yet defined).
- Observation Reports
  - -- Statements in the algebra.
- Updating Mechanism
  - -- Via Bayes' Theorem, calculate posterior probabilities based on prior and conditional probabilities. In symbolic form,

$$P_r(S_i/E_j) = \frac{P_r(S_i)\ P_r(E_j/S_i)}{\sum_k P_r(S_k)\ P_r(E_j/S_k)},$$

where Pr is the r^{th} member of the convex set of probability functions.

- Decision Mechanism
  - -- Upper and lower probabilities for some statement in the algebra can be taken from the convex set of $P_r(S_i/E_j)$, but no general procedure exists to handle these in a utility function. Presumably the decision indicated is the one that maximizes this utility using the convex set of $P_r(S_i/E_j)$.

3. Approach 3: Dempster-Shafer

a. Background Elements

The background in this approach, like the first two approaches, consists of three elements: (1) an algebra of statements, (2) a mass function defined over this algebra, and (3) a utility function defined over the same algebra. The algebra defines the domain of discourse, the mass function assigns degrees of belief to elements of the domain, and the utility function provides a means of reaching decisions in the domain. It should be noted that the utility function has received little attention in this approach, but will be required in practical applications.

The Dempster-Shafer approach also uses an LT algebra consisting of base elements, operators, and propositions entailed by application of the operators to the base elements. The base elements are again assumed to be mutually exclusive. We shall continue to refer to the mutually exclusive elements as base elements and to the legal statements as atoms.

The mass function serves as the basic vehicle for assignment and manipulation of degrees of belief. Mass is distributed across the set of subsets of the elements of the domain of discourse, that is, over the set S of (2 exp 2^n) propositions constructed from the 2^n atoms that were in turn constructed from the n base elements.

The mass function m_1 for subset A_i of S has the following properties (S4):

$$m_1(A_i) \text{ is a real number on } [0,1]$$

$$m_1(\text{null set}) = 0$$

$$\sum_i m_1(A_i) = 1 \quad .$$

The value of $M_1(f_i)$ is taken to be the weight of belief that is ascribed just to f_i. The f_i for which $M_1(f_i)$ is nonzero are called focal elements of M_1. Since S is itself a member of S, $M_1(S)$ describes the weight of belief unassigned to any smaller subset of S; this may be termed the uncertainty.

This approach provides two measures of belief state for a given proposition Q: support (SPT) and plausibility (PLS). They are calculated as follows (B2, S4):

$$SPT_1(Q) = \sum_{f_i \subseteq Q} M_1(f_i) \quad (3.1),$$

$$PLS_1(Q) = 1 - \sum_{f_i \subseteq \sim Q} M_1(f_i) \quad (3.2),$$

$$= 1 - SPT_1(\sim Q) \quad (3.3).$$

The support for Q is thus the sum of the mass attributed to all subsets of Q, while the plausibility of Q is one minus the support for the negation of Q. The plausibility can also be expressed as the sum of the mass attributed to all subsets of S that contain some element of Q. It follows that the plausibility of Q is always greater than or equal to the support for Q.

The belief state concerning Q can be written as an interval using SPT(Q) as the lower endpoint and PLS(Q) as the upper. Some authors describe this as an interval-valued probability on Q (D2). Kyburg has shown (K4) that closed convex sets of classical probability functions can represent belief states in a fashion that includes the mass-function representation as a special case.

The background also contains means of translating observation reports into mass functions. One method is that of a mass-function distribution; this distribution provides a normalized measure of the mass to be assigned to each element of the domain in the event of each possible observation.

Figure II-4 shows a set of such distributions in schematic form. These are clearly analogous to the class-conditional functions of standard probability theory.

b. Observation Reports

Observation reports, at least to the extent that they are expected to mesh with mass-function distribution, consist of statements like the following:

> S1 - "The brightness of object X is between 1.2 and 1.6."
>
> S2 - "Object X is surrounded by between 2 and 6 objects of similar brightness."

Proponents of this approach assert that it is not limited to the handling of data based upon observational statistics, so reports might also consist of statements that embody knowledge that is not necessarily based upon statistical data. An example of such a report is:

> S3 - "In region Y, the expectation of encountering an object of class C_1 is much higher than that of any other class."

In any case the intent of the approach is that observation reports determine mass functions via mass-function distributions.

It is to be noted that each type of observation report is taken to generate a separate mass function. This presents no problem as long as it is clear that the evidential impact of a given report is properly assigned to some subset of the domain of discourse. However, how this proper assignment is to be guaranteed is not a trivial matter. For example, suppose we were to receive the following report:

> S4 - "The R-brightness of object X is between 102 and 108."

If our domain of discourse were constructed to deal only with reports on brightness of an object, the brightness of its nearest neighbors, and the expectation of encountering certain classes of objects in certain regions,

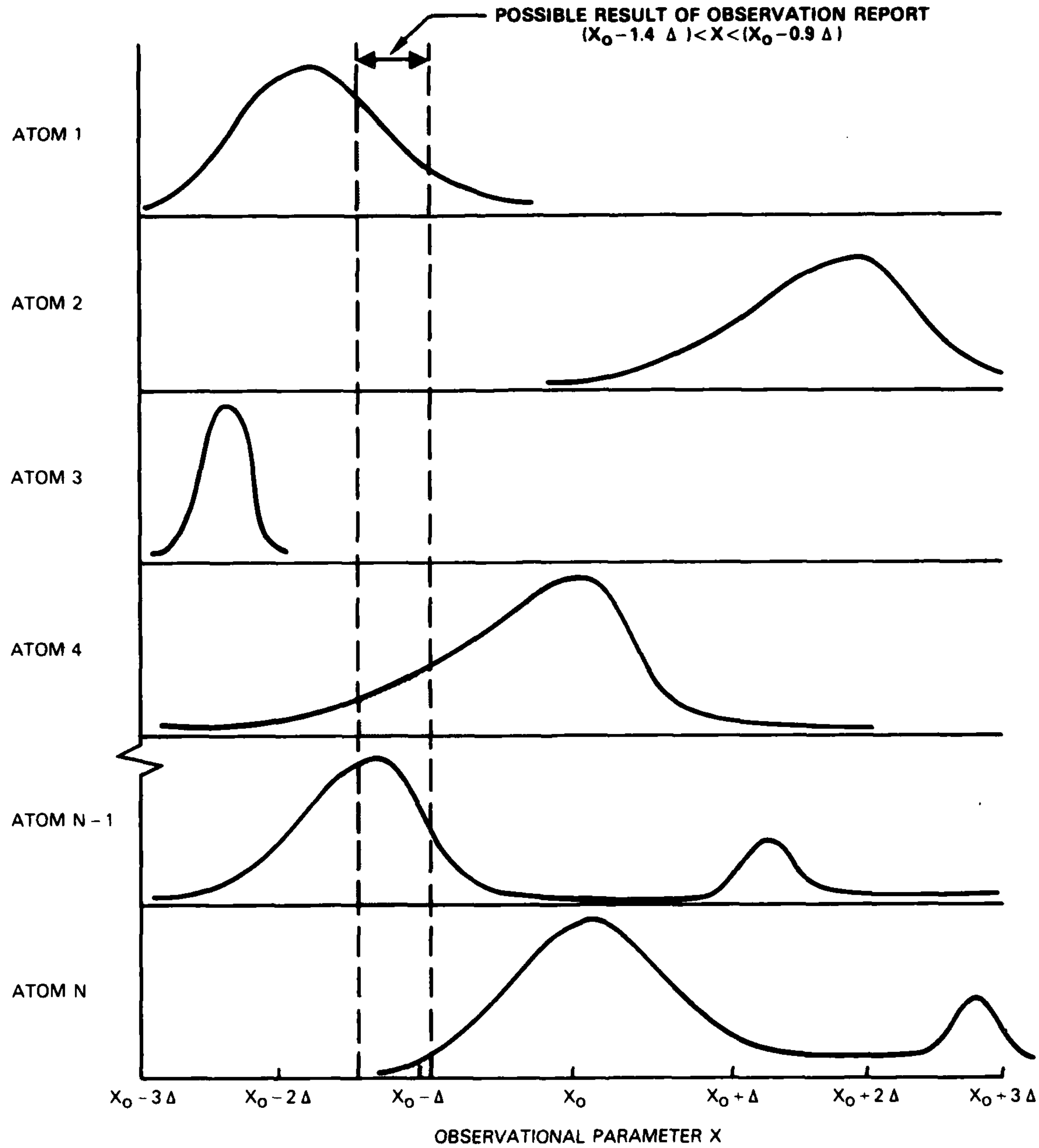


FIGURE II-4.
SCHEMATIC MASS-FUNCTION DISTRIBUTIONS IN THE DEMPSTER-SHAFER APPROACH

then we would not know what to do with this report on R-brightness. Clearly, the domain of discourse would require restructuring if we intend to make use of such reports.

This raises the question of what rules we should use in structuring the domain of discourse. We want to ensure the inclusion of subsets that can serve as recipients of mass from each and every observation report that will be received in performance of a given task. A general theory of the development of such rules is an area of much current research and is beyond the scope of this report. However, several practical insights on this topic will emerge from our discussion of sample image-analysis tasks in Chapter III.

c. <u>Updating Mechanism</u>

Suppose that we have received two observation reports that have individually engendered mass functions M_1 and M_2. We combine M_1 and M_2 to form a new mass function, M_{12}, defined over subsets of the domain of discourse (S4). In symbolic form,

$$M_{12}(f_k) = \frac{\sum' M_1(f_i)\, M_2(f_j)}{1 - \sum'' M_1(f_i)\, M_2(f_j)} \qquad (3.4),$$

where the first summation, $\sum'$, is over all f_i and f_j such that $(f_i \wedge f_j) = f_k$, while the second summation, $\sum''$, is over all f_i and f_j such that $(f_i \wedge f_j) = \text{null}$.

The updating procedure first assumes that a current mass function, M_1, is available. Then it assumes that a new mass function, M_2, has been presented. Finally it combines M_1 and M_2 to form M_{12}, and uses it as the current mass function should other new mass functions be presented. Figure II-5 gives an overview of this process.

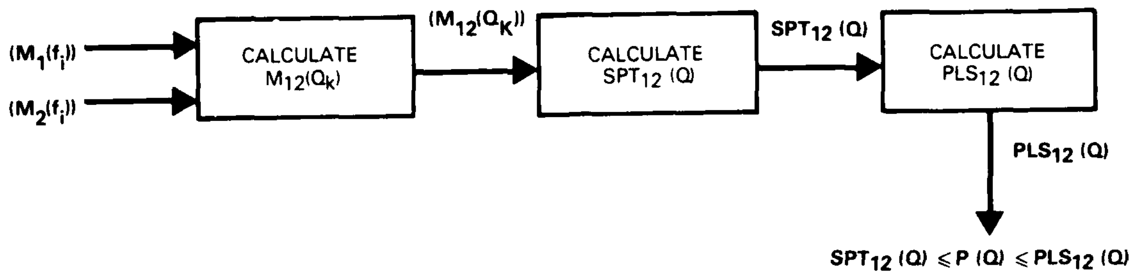


FIGURE II-5.
OVERVIEW OF DEMPSTER-SHAFER UPDATING

d. Decision Mechanism

The type of decision mechanism compatible with the Dempster-Shafer approach is not currently known. Support and plausibility functions for each statement in the domain of discourse can be calculated based upon the current mass function. These may be used as upper and lower bounds upon the probability of each statement, but there is as yet no accepted, general mechanism for decision-making based upon these bounds.

If we attempt a construction parallel to the classical and convex Bayesian approaches, the difficulties become apparent. As before, we formulate the expected loss of the i^{th} action as follows:

$$EL_i = \sum_j (l_{ij} * PCUR_j) \quad (3.5),$$

where the summation is over the j states of nature. If we construe $PCUR_j$ as an interval bounded by $SPT(Q_j)$ and $PLS(Q_j)$, then we might say that EL_i lies between $ELMIN_i$ and $ELMAX_i$, where

$$ELMIN_i = \sum_j (l_{ij} * SPT(Q_j)) \quad (3.6),$$

$$ELMAX_i = \sum_j (l_{ij} * PLS(Q_j)) \quad (3.7).$$

For different actions, these intervals will, in general, overlap.

If the decision function is simply to chose, whenever a decision is required, the action that gives the minimum value of EL_i, it is not clear in this case how to determine the action that conforms to this decision rule. Some simplification may be obtained by partitioning the set of EL-intervals into those that might include the minimum and those that will not, but there remains the problem of choosing the appropriate interval from the candidate subset. An analogue of mixed strategies may be useful, but this remains an open question for research.

Table III-3 provides a brief summary of this approach. It may be compared with similar tables for the other approaches.

4. Approach 4: Kyburg

a. Background Elements

The background in this approach, like others already discussed, consists of three elements: (1) an algebra of statements, (2) a probability function defined over this algebra, and (3) a utility or loss function defined over the same algebra. As before, the algebra defines the domain of discourse, the probability function assigns degrees of belief to elements of the domain, and the utility function provides a means of reaching decisions in the domain.

The Kyburg approach also uses an LT algebra consisting of base elements, operators, and propositions entailed by application of the operators to the base elements. As before, the base elements are assumed to be mutually exclusive, so application of the disjunctive operator expands them into the set of all possible legal statements about the domain of discourse.

TABLE II-3

SUMMARY OF DEMPSTER-SHAFER APPROACH

- Background Elements

  -- Algebra of statements, AS-1

  -- Mass function M1 defined over subsets of AS-1

- Observation Reports

  -- Mass function M2 defined over the subsets of another algebra of statements, AS-2.

  -- This mass function can sometimes be decomposed into mass supporting statements in AS-1 and mass supporting the set of all subsets of AS-1.

- Updating Mechanism

  -- Combine M_1 and M_2 to form a new mass function, M_{12}, defined over subsets of AS-1. In symbolic form,

$$M_{12}(f_k) = \frac{\sum' M_1(f_i)\, M_2(f_j)}{1 - \sum'' M_1(f_i)\, M_2(f_j)} ,$$

  where the first summation, $\sum'$, is over all f_i and f_j such that $(f_i \wedge f_j) = f_k$, while the second summation, '', is over all f_i and f_j such that $(f_i \wedge f_j) = \text{null}$.

  -- This formula is used as follows:

    •• Assume that a current mass function, M_1, is available.

    •• Assume that a new mass function, M_2, has been presented.

    •• Combine M_1 and M_2 to form M_{12}, and use it as the current mass function should other new mass functions presented.

- Decision Mechanism

  -- Support and plausibility functions for each statement in AS-1 can be calculated based upon the current mass function. These may be used as upper and lower bounds upon the probability of each statement, but there is as yet no general mechanism for decision-making based upon these bounds.

Direct inference refers to the manner in which knowledge of chances (or frequencies, or objective probabilities) influences belief states about the outcomes of trials involving chance setups. In Kyburg's approach, some portion of the algebra of statements has the status of a body of knowledge containing statements about relative frequencies of occurrence of various characteristics in various classes.

Kyburg offers a principle of direct inference that allows the assignment of precise or imprecise p-values to hypotheses based upon knowledge of frequencies without requiring the assignment of precise prior p-values. Adoption of his principle of direct inference may provide support for the use of Fisherian fiducial inference.

The concept of probability embraced by this approach is epistemological. This means that probability is actually a descriptor of credibility relative to some body of knowledge. In addition, the p-value used in this approach is an interval on [0,1].

We will use a concrete example as the basis of our discussion of the Kyburg approach. The body of knowledge is taken to consist of the following statements:

- The fraction of members of class C_1 that have property P lies in the interval $[L_1,U_1]$.
- The fraction of members of class C_2 that have property P lies in the interval $[L_2,U_2]$.
- The fraction of members of class C_{12} that have property P lies in the interval $[L_{12},U_{12}]$, where class C_{12} is the intersection of classes C_1 and C_2.

The hypothesis of interest is that an item selected from class C_{12} has property P.

In order to show how to apply Kyburg's direct-inference principle, we require a means of determining which evidence is relevant to a given statistical hypothesis and which is not. We shall use two terms, K-relevance and K-irrelevance, to this end.

K-irrelevance refers to a mandatory lack of impact of a given piece of information on our deliberations concerning the credibility of a certain statistical hypothesis. The information concerning C_2 in the body of knowledge is K-irrelevant if and only if the following conditions are fulfilled:

- The current body of knowledge implies that $[L_1, U_1]$ is either a subinterval of, or identical to, $[L_{12}, U_{12}]$.
- The current body of knowledge implies that $[L_1, U_1]$ is either a subinterval of, or identical to, $[L_2, U_2]$.

In our example, if the information concerning C_2 is K-irrelevant, then the information concerning C_1 is the total information K-relevant to the hypothesis.

K-irrelevance is thus a formal criterion that tells us whether or not knowledge of a specific relative frequency should influence our degree of belief that a member of C_{12} has property P. For example, we might know that:

- The fraction of Swedes who are Protestant lies in [a, b].
- The fraction of visitors in Lourdes who are Swedish lies in [c, d].
- The fraction of Swedish visitors to Lourdes who are also Protestant lies in [e, f].

We might then be interested in finding the appropriate degree of belief to attach to the hypothesis that a particular person is a Protestant, given that we know that he is a Swedish visitor to Lourdes. Intuitively, we know that the values of a, b, c, d, e and f will influence this degree of belief. K-irrelevance formalizes this process.

Kyburg's principle of direct inference has a simple form, once the criterion of K-irrelevance has been applied to the body of knowledge. It states that, if the information concerning C_1 is the total information K-relevant to the hypothesis, then the degree of belief to be assigned to the hypothesis is just the interval $[L_1, U_1]$.

The upshot of this process is that the Kyburg approach recommends, in many cases, that different intervals of degrees of belief be embraced. This has the consequence that the evolution of p-values as evidence accumulates follows a different trajectory through the space of belief states. That this different trajectory may have important practical impact seems reasonable, but remains to be demonstrated in a systematic fashion.

b. Observation Reports

Observation reports in this approach can again be construed as statements in the algebra. When coupled with appropriate knowledge of relative frequencies, they assign new p-values to elements of the algebra. Such assignments explicity refer to interval-valued p-functions.

There is just one form of report in this approach:

- Each observation report consists of the identification of the class or classes to which the observed object belongs. Knowledge of relative frequencies then determines how to assign two p-values in the interval [0,1] to some element in the algebra. These are construed as lower and upper p-values for the element.

An updating mechanism to handle such reports is available, but, as in other approaches using interval-valued p-functions, the decision mechanism is an area of ongoing research.

c. Updating Mechanism

The Kyburg approach mandates the use of a unique updating mechanism: the principle of direct inference discussed above. In some special cases, this gives results that can be obtained from Bayes' Theorem. Legal applications of the theorem are possible when the body of knowledge contains statements like:

S1 - The fraction of members of class C_i that have property P lies in the interval $[L_i, U_i]$.

These statements must be based on knowledge of relative frequencies of occurrence of properties in real sets of objects or events. Furthermore, for each hypothesis being considered, the appropriate sorting of the elements of the body of knowledge into those that are K-relevant and those that are not must take place.

Given these conditions, the application of Bayes' Theorem proceeds in a manner analogous to that of the convex Bayes approach. The principal difference is that the size of the convex set of p-functions calculated using intervals alone usually will be smaller than the set of functions considered in the explicitly convex approach.

d. Decision Mechanism

The Kyburg approach offers interval-valued p-functions. As has been discussed above for both the convex Bayes and Dempster-Shafer approaches, there is currently no general decision mechanism available for interval-valued p-functions.

We close this section with a summary of the Kyburg approach (Table II-4). This may be compared with similar summaries for the other approaches.

5. Approach 5: Neyman-Pearson

a. Background Elements

We focus here on the body of theory developed by J. Neyman and E. Pearson to deal with the testing of hypotheses. We shall discuss several important features of this approach: first, the nature of the hypotheses being assessed; second, the means by which evidential weights are developed for each hypothesis; and third, the means by which the decisions are made on the basis of the weights.

The hypotheses treated by this approach are statistical in nature. That is, they concern the behaviour of observable random variables. Some authors, however, do not explicitly restrict the hypotheses in this fashion.

TABLE II-4
SUMMARY OF KYBURG APPROACH

- Background Elements

  -- Algebra of statements

  -- Body of knowledge defined over subsets of the algebra

  -- Criteria of relevance concerning statistical hypotheses.

- Observation Reports

  -- Each observation report consists of the identification of the class or classes to which the observed object belongs. Knowledge of relative frequencies then determines how to assign two p-values in the interval [0,1] to some element in the algebra. These are construed as lower and upper p-values for the element.

- Updating Mechanism

  -- Apply criteria of relevance to elements of the body of knowledge

  -- Apply principle of direct inference as appropriate

  -- Bayes' Theorem may be applied in certain cases involving knowledge of relative frequencies of occurrence of properties in real sets of objects or events.

- Decision Mechanism

  -- Interval-valued p-functions are derived for statements in the algebra. These may be used as upper and lower bounds upon the probability of each statement, but there is as yet no general mechanism for decision-making based upon these bounds.

We may divide this approach into three segments. In hypothesis testing, we reason from observations to conclusions about whether the state of nature falls into one of two categories, and then chose one of two actions. In estimation, we reason from observations to conclusions about whether the state of nature falls into one of n categories ($n > 2$), and then chose one of n actions that lie in one-to-one correspondence with the states of nature. In confidence-interval generation, we reason from observations to conclusions about the set of categories in which the state of nature falls, and then chose one of several actions. The size of the set of categories is determined by the confidence level, L, such that, presumably, the set is larger the closer L is to unity. The size of the set is also determined by the nature, the number, and possibly the sequence of the observations in a way that must be specified for each application.

The major elements of the background are an algebra of statements and set of confidence intervals. The algebra of statements defines the domain of discourse in a manner similar to that of the approaches discussed above. The confidence intervals provide the linkage between observation reports and the correlate of p-values to be assigned to statements in the algebra. These values will, in general, also be intervals.

The confidence intervals are to be constructed as follows:

- For a possible value, x_0, of some parameter, X, in the domain of discourse, construct the set of values of some other parameter, Z, that constitute a test of significance level (1-L) that X has the value x_0. Call this set of Z values the acceptance set for x_0 on Z and denote it by AS(z: $x = x_0$).
- Repeat this procedure for all other possible values of X to obtain the set of all acceptance sets for x_i on Z, AS(z: $x = x_i$) . This set of acceptance sets can be visualized as creating envelope in the X-Z plane as shown in Figure II-6.
- From this envelope find the set of values of X that would not be rejected by observation of value z_0 for Z. This is the confidence interval of level L for X on z_0, denoted by CI(x: $z = z_0$).

The confidence coefficient that the interval so constructed contains the actual value of x is then taken to be L. We may, with caution, correlate the confidence coefficient with the p-value discussed in other

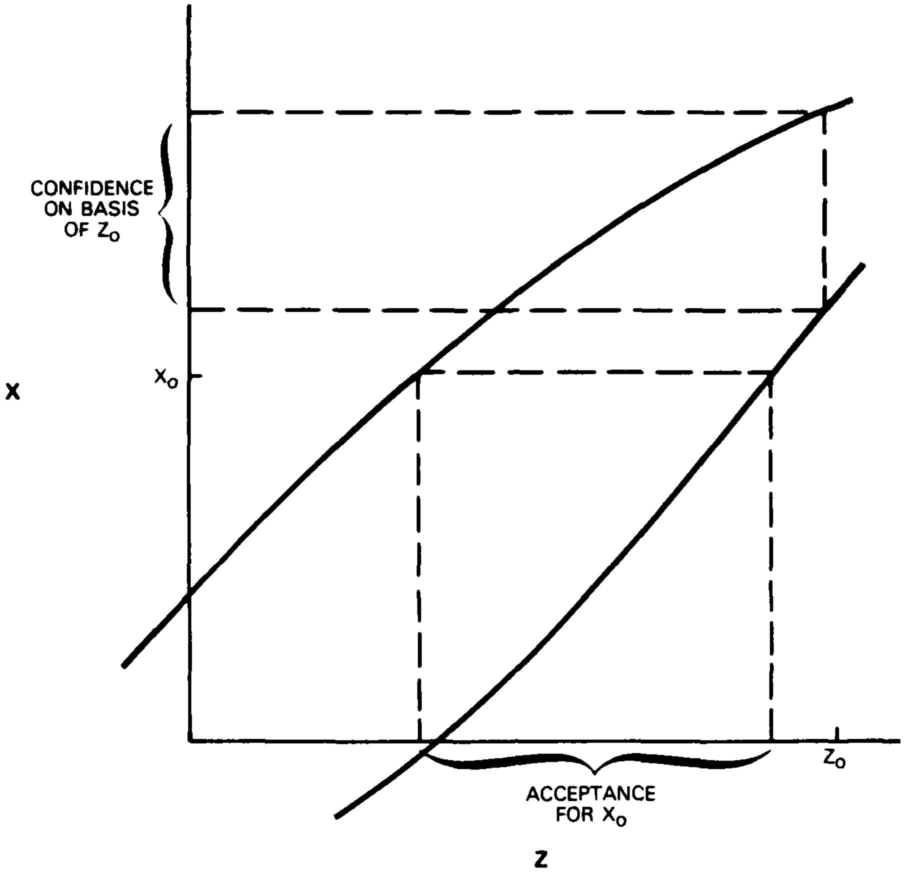


**FIGURE II-6.
CONFIDENCE-INTERVAL CONSTRUCTION**

The confidence coefficient that the interval so constructed contains the actual value of x is then taken to be L. We may, with caution, correlate the confidence coefficient with the p-value discussed in other approaches. The set of all confidence intervals of level L for X on z_i may be termed the confidence set for X on Z, denoted by CI(x: z = z_0).

Other parameters in the domain must also be addressed. That is, one must: (1) repeat these steps for all other possible parameters that can constitute tests for X, and (2) repeat these steps for all other parameters in the domain of discourse that stand in the relation of testing as do X and Z.

Examples of pairs of X-like and Z-like parameters might include:

- Z - The position of the pointer on an analog voltmeter

  X - The steady-state value of the voltage in a certain circuit
- Z - The brightness of a certain pixel

  X - The strength associated with the assignment of a pixel to a certain class.

Clearly, the nature and number of such pairs will influence the complexity of the updating process.

b. Observation Reports

Observation reports are statements in the algebra. They refer to parameters like Z that have been identified during construction of the confidence sets.

There are two possible interpretations of the content of observation reports that could be addressed using this approach:

- Each report consists of the assignment of a single p-value of 1.0 to some element in the algebra of statements that corresponds to a single value of the observation parameter
- Each report consists of the assignment of a single p-value of 1.0 to some element in the algebra of statements that corresponds to a range of values of the observation parameter.

It was apparently the first of these interpretations that was of concern to the developers of this approach.

The second interpretation may offer a means of dealing with uncertainty in observation reports. For example, suppose we are dealing with reports similar to the following:

R1 - "The degree to which this feature resembles an edge lies between 8 and 9 (on a scale of 10)."

Assuming that the confidence sets have been constructed appropriately, then the ends of the reported interval, z_1 and z_2, could be used to locate two different confidence intervals, $CI_1(x\colon z = z_1)$ and $CI_2(x\colon z = z_2)$. The confidence interval to be used with the given operation might then be taken to be that defined by the minimum and maximum values of x found in the interval obtained by combining CI_1 and CI_2.

It is not clear, however, whether it is possible to carry through the construction of confidence intervals using the second interpretation. Clearly, the envelope must be of the simple shape depicted in Figure II-6 for this interpretation to be applicable. More importantly, the proper construction of tests of significance level (1-L) under these conditions remains a question for further research.

c. Updating Mechanism

There are two updating mechanisms available in this approach. First, updating can occur as a result of a report concerning additional observations of the same Z-like parameter. Second, updating can occur as a result of a report concerning observations of other Z-like parameters.

With regard to the first mechanism, the confidence set described in section (a) can be construed as a family of nested curves that define a region in the X-Z plane that grows smaller as additional observations are made. This means that an observed value of Z determines a smaller confidence interval when more observations, rather than fewer, have been made.

In this case the updating proceeds as follows:

- Receive observation report that gives the value of Z
- Determine the number of observations of this type and use the curves in the confidence set corresponding to this number
- From these curves, determine the limits on X to confidence level L.

It is clear that this mechanism can be applied to certain cases of evidential reasoning, particularly those that involve sampling from populations that can be regarded as examples of well-defined statistical distributions. The degree to which the mechanism can be applied to situations that do not entail such distributions is not clear.

The second updating mechanism operates in a fashion similar to the first, except that the receipt of reports on two Z-like parameters now requires that the confidence intervals from these two parameters be integrated. How this is to be done is, to the author's knowledge, not addressed by practitioners of this approach. Perhaps the answer lies in the construction of the algebra of statements so that all relevant combinations of Z-like parameters are represented by appropriate confidence sets.

It should be noted, for both mechanisms, that it is not the p-value of X that changes as new reports are received; this p-value is fixed by the choice of L. The change induced by new reports is in the limits of the confidence interval that is believed, with a p-value of L, to contain the actual value of X.

d. Decision Mechanism

The Neyman-Pearson approach, as here described, provides us with p-functions of value L on statements concerning intervals within which parameters may fall. If the algebra and the actions have been constructed with this in mind, it may be possible to map preferred actions to intervals in the X domain. In that case, the choice of action is clear if the confidence interval on X coincides with one of the action intervals. However, the intervals will not coincide in general, and the appropriate choice of action is problematic. Weighting of the loss factors by measures of interval overlap (in addition to the confidence level) may be feasible, but will not be pursued here.

We conclude this section with Table II-5. This presents a brief summary of the Neyman-Pearson approach.

TABLE II-5

SUMMARY OF NEYMAN-PEARSON APPROACH

- Background Elements
  - -- Algebra of statements
  - -- Confidence intervals of level L defined over subsets of the algebra.
- Observation Reports
  - -- Each report consists of the assignment of a single p-value of 1.0 to some element in the algebra of statements that corresponds to a single value of the observation parameter.
  - -- In some cases, the element of the algebra may correspond to a range of values of the observation parameter, rather than a single value.
- Updating Mechanism
  - -- For additional observations of the same Z-like parameter:
    - •• Receive observation report that gives the value of Z.
    - •• Determine the number of observations of this type and use the curves in the confidence set corresponding to this number.
    - •• From these curves, determine the limits on X to confidence level L.
  - -- For observations of different Z-like parameters, the procedure is presumably similar if the algebra of statements has been constructed so that all relevant combinations of Z-like parameters are represented by appropriate confidence sets.
  - -- The change induced by new reports is in the limits of the confidence interval that is believed, with a p-value of L, to contain the actual value of X.
- Decision Mechanism
  - -- Single-valued p-functions are derived for statements in the algebra concerning intervals that are believed to contain the actual value of X. If actions have been defined in terms of intervals of X, the preferred action will be clear if action and confidence intervals coincide. If they do not, weighting based upon measures of interval overlap may be feasible.

6. Approach 6: Possibility

a. Background Elements

The background in this approach consists of three elements: (1) an algebra of statements, (2) degree-of-membership functions defined over this algebra, and (3) a set of fuzzy decision functions defined over the same algebra. The algebra defines the domain of discourse, the membership function assigns degrees of membership to elements of the domain, and the decision functions provide a means of reaching decisions in the domain.

The degree-of-membership function is defined in terms of a fuzzy set. Such a set is made up of ordered pairs that assign a degree of membership in the fuzzy set to each value of a given characteristic. There is one specific characteristic associated with each set. The fuzzy set is then denoted by

$$A = \{x_i | p_i\},$$

where x_i is the i^{th} value of the characteristic and p_i is the degree of membership of x_i in the fuzzy set.

In Zadeh's fuzzy logic, p-values obey the following axioms (G1, Z2):

$$0 < p(x) < 1$$

$$p(\sim x) = 1 - p(x)$$

$$p(x \wedge y) = \min[p(x), p(y)]$$

$$p(x \vee y) = \max[p(x), p(y)]$$

$$p(x \rightarrow y) = \min \{1, [1 - p(x) + p(y)]\}$$

$$p(x = y) = \min \{[1 - p(x) + p(y)], [1 + p(x) - p(y)]\}.$$

The last four axioms constitute strong truth functionality (G1, p.55).

The background also contains definitions of fuzzy predicates appropriate to the domain of discourse. These are statements that establish the degree of membership in a fuzzy set as a function of some characteristic of an object. For example, the following statements define the fuzzy predicate for low reflectance (LR) based on values of the reflectance:

S1 - "A reflectance of 0.0 is low with degree 1.0."
S2 - "A reflectance of 1.0 is low with degree 0.8."
S3 - "A reflectance of 2.0 is low with degree 0.4."
S4 - "A reflectance of 3.0 is low with degree 0.2."
S5 - "A reflectance of 4.0 is low with degree 0.1."

The fuzzy set would then be represented by

$$LR = \{0|1.0, 1|0.8, 2|0.4, 3|0.2, 4|0.1\}.$$

b. Observation Reports

Observation reports in the possibility approach provide the raw material for assignment of degrees of membership . That is, they are statements that establish the degree-of-membership value of a characteristic of an object. Application of the membership function then determines the degree of membership of that object in the fuzzy set.

For example, suppose that statements S1 through S5 above define the low-reflectance fuzzy set. The following report would establish the degree of membership of Z in the class of low-reflectance objects as 0.8:

S6 - "The reflectance of object Z is 1.0."

Note that the algebra is here construed as operating on the LR predicate, not on the reflectance values. The number of fuzzy predicates is an important determinant of the size of the algebra of statements.

c. Updating Mechanism

The possibility approach combines evidence in the following fashion. Suppose we desire to classify a certain object into one of n classes,

$c_1 \ldots c_n$. Based upon evidence E_1, we develop membership functions p_{11} through p_{1n} to form the fuzzy set

$$A_1 = \{ c_1 | p_{11}, c_2 | p_{12}, \ldots, c_n | p_{1n} \} .$$

Similarly, for evidence E_2,

$$A_2 = \{ c_1 | p_{21}, c_2 | p_{22}, \ldots, c_n | p_{2n} \} .$$

We combine k sets of evidence to obtain

$$B(k) = \{ c_1 | p(k)_1, c_2 | p(k)_2, \ldots, c_n | p(k)_n \} ,$$

where the $p(k)_1, \ldots, p(k)_n$ are integrated membership functions for each of the n classes. These are obtained from

$$p(k)_j = D_{xxx}(p_{1j}, p_{2j}, \ldots, p_{kj}) \qquad (6.1),$$

where D_{xxx} is one of several alternative fuzzy decision functions:

$$D_{int}(p_{1j}, \ldots, p_{kj}) = MIN(p_{1j}, \ldots, p_{kj}) , \qquad (6.2),$$

$$D_{pro}(p_{1j}, \ldots, p_{kj}) = \prod_{i=1}^{k} p_{ij} \qquad (6.3),$$

$$D_{con}(p_{1j}, \ldots, p_{kj}) = \sum_{i=1}^{k} a_{ij} p_{ij} , \quad (\sum_{i=1}^{k} a_{ij} = 1) \qquad (6.4).$$

Use of D_{int} suggests that E_1 and E_2 interact in a a more or less independent fashion, and that the presence of a smaller p-value should be preserved. Use of D_{pro} suggests that E_1 and E_2 interact like identical, independent trials, so that repetitive observations cause marked changes in relative values of membership. Use of D_{con} suggests that E_1 and E_2 interact in a reinforcing fashion, so that membership is intermediate between the two input values.

d. Decision Mechanism

The decision mechanism in the possibility approach is based upon the concepts of the fuzzy goal and the fuzzy constraint. The essential idea is that decisions are determined by the confluence of goals and constraints, and that all three are expressible as fuzzy sets (B4).

For example, suppose that the domain of discourse has been constructed to allow expression of goals and constraints in the same algebra of statements, $\{S_1, \ldots, S_n\}$. Then the expression of goals would be embodied in the fuzzy set

$$G = \{S_1|m_{1g}, S_2|m_{2g}, \ldots, S_n|m_{ng}\} \qquad (6.5).$$

The constraints would be embodied in a similar fuzzy set

$$C = \{S_1|m_{1c}, S_2|m_{2c}, \ldots, S_n|m_{nc}\} \qquad (6.6).$$

The confluence of goals and constraints is expressed by the fuzzy set

$$DEC(G,C) = \{S_1|m_{1gc}, S_2|m_{2gc}, \ldots, S_n|m_{ngc}\} \qquad (6.7),$$

where the $m_{1gc}, \ldots, m_{ngc}$ are integrated membership functions for the goals and constraints. These are obtained from

$$m_{jgc} = D_{xxx}(m_{jg}, m_{jc}) \qquad (6.8),$$

where D_{xxx} is one of the fuzzy decision functions discussed above.

The alternative decision functions now have only two arguments, but function as follows:

- D_{int} makes a component of the decision space reflect only the lesser of the goal or constraint values
- D_{pro} makes a component of the decision space reflect the product of the goal and constraint values
- D_{con} makes a component of the decision space reflect the weighted average of the goal and constraint values.

General criteria for the choice among these decision rules have not been developed.

Once the confluence set, DEC(G,C), has been constructed, the question remains as to which decision is indicated. Several procedures are followed in the literature (B4, M2). The most notable are: (1) choice of the action having the greatest DEC degree of membership, (2) choice of an action that is a combination of all actions weighted according to their DEC degrees of membership, and (3) choice of an action that is an equal mixture of the two actions having the minimum and maximum DEC degrees of membership. General criteria for the choice among these approaches have not been developed.

We conclude this section with a brief summary of the possibility approach (Table II-6).

TABLE II-6
SUMMARY OF POSSIBILITY APPROACH

- Background Elements
  - Algebra of statements
  - Fuzzy predicates defining degrees of membership on subsets of the algebra.

- Observation Reports
  - Reports establish the value of a characteristic of an object.
  - Application of the membership function then determines the degree of membership of that object in the appropriate fuzzy predicate set.

- Updating Mechanism
  - Membership functions are combined via one of three alternative decision functions:

$$D_{int}(p_{1j}, \ldots, p_{kj}) = \mathrm{MIN}(p_{1j}, \ldots, p_{kj}) \quad ,$$

$$D_{pro}(p_{1j}, \ldots, p_{kj}) = \prod_{i=1}^{k} p_{ij} \quad ,$$

$$D_{con}(p_{1j}, \ldots, p_{kj}) = \sum_{i=1}^{k} a_{ij} p_{ij} \quad , \quad \left(\sum_{i=1}^{k} a_{ij} = 1 \right) \quad .$$

- Decision Mechanism
  - Decisions determined by the confluence of goals and constraints expressed as fuzzy sets
  - The confluence is also a fuzzy set obtained from the goal and constraint sets via application of the fuzzy decision functions
  - The decision is ordinarily indicated by the maximum value of the confluence set, but other indicators are sometimes used.

D. COMPARATIVE ANALYSIS OF THE APPROACHES

1. General

From the discussion in Section C, we can identify several important similarities and differences:

- The structure, but not necessarily the content, of the algebra of statements is similar across the approaches.
- All approaches, with the possible exception of the Neyman-Pearson and Possibility approaches, depend on formulation of a loss function to arrive at decisions.
- Structures given to belief states are significantly different. They may be points, intervals, convex sets, or fuzzy sets.
- Components of the updating mechanism differ significantly, but the approaches fall into three major categories: those that use Bayes' Theorem exclusively (classical and convex Bayes), those that allow its use under certain conditions (Kyburg) or use a derivative form (Dempster-Shafer), and those that do not use it at all (Neyman-Pearson and Possibility).
- The approaches differ in their treatment of confirmational conditionalization, direct inference, and committment to numerically precise priors (L2, p.369). These are discussed below.
- Components of the decision mechanism also differ. Extension of the expected-loss technique from the case of point-valued belief states to the cases of interval-valued or convex-set belief states may be possible, but precisely how this is to be done remains an open question. The fuzzy decision rules operating on the confluence of fuzzy goals and constraints appear to be unique.

We elaborate on several of these points below.

The belief states are characterized in the classical Bayes and Neyman-Pearson approaches as single points on the interval [0,1]. In the Kyburg, Neyman-Pearson, and Dempster-Shafer approaches they are construed as subintervals of [0,1]. In the convex Bayes approach they are construed as convex sets of functions in a space of (n-1) dimensions, where n is the number of base elements in the domain of discourse.

The fuzzy-set approach is purportedly different in that it is concerned with grades of membership rather than degrees of belief. However, since it also computes numbers that are essentially p-values, the distinction between grades of membership and degrees of belief is less important than it might seem at first glance. The real distinction lies in the manner in which p-values are combined. It appears that p-values can be either single points or sub-intervals of [0,1].

The approaches differ in the exercise of both direct and inverse inference. Direct inference refers to the manner in which knowledge of chances (or frequencies, or objective probabilities) influences belief states about the outcomes of trials involving chance setups. There is a group of direct-inference principles endorsed by various authors (L2, pp.54, 250f), but they will not be discussed here. Inverse inference refers to the manner in which knowledge of the outcomes of trials involving chance setups influences belief states about chances (or frequencies, or objective probabilities). These are belief states concerning rival statistical hypotheses (L2).

The key point is that rules governing direct inference are not identical to rules governing inverse inference. The exposition, clarification, and comparison of these rules is beyond the scope of the current effort, but we may note that the convex Bayes approach differs from the Kyburg approach. The former allows direct inference from beliefs, frequencies, or knowledge of chances, while the latter prefers direct inference only from knowledge of relative frequencies (L2, p.393).

The approaches also differ in terms of confirmational conditionalization. This refers to the manner in which belief states change when there is a transition from an old to a new body of knowledge. In its simplest form, confirmational conditionalization requires that the same rule that generated the belief state based on the old body of knowledge be used in the same way to generate the belief state based on the new body of knowledge. The convex Bayes approach embraces confirmational conditionalization, while the Kyburg approach does not. The Dempster-Shafer approach apparently relies on its own form of confirmational conditionalization, called D-conditionalization.

2. Points of Correspondence

The most salient point of correspondence is the dependence, in each approach, upon an algebra of statements that sharply defines the domain of discourse. This algebra is a fixed framework within which all observations are interpreted, all updating occurs, and all decisions are made. None of the approaches discussed here addresses the issue of a dynamic algebra that adapts to changing real-world conditions.

A second point of correspondence is dependence upon definition of a loss function. This is explicit in all but the Neyman-Pearson and Possibility approaches, and may even be implicit in these. Certainly the construction of appropriate loss functions is not a trivial matter, nor is there an iron-clad general theory of utility functions available to guide us. A third point of correspondence has recently been elucidated by Kyburg. He has shown (K4) that the probability intervals resulting from application of the Dempster-Shafer updating mechanism are included in the intervals resulting from Bayesian updating on the same evidence. He has also shown that closed convex sets of classical probability functions provide a representation of belief states that includes the mass-function representation as a special case.

3. Points of Incommensurability

A key point of incommensurability is concerned with whether Bayes' Theorem can be used to calculate precise posterior p-values concerning the outcomes of trials from prior p-values concerning such outcomes. The Kyburg view is that this is acceptable only when precise or imprecise priors based upon knowledge of frequencies are available via direct inference. On this view, there are circumstances in which no legitimate priors other than the entire interval [0, 1] are available; consequently, there are circumstances in which Bayes' Theorem cannot be applied.

A second point of incommensurability is related to the structure of belief states. In a certain sense, higher-order representations can subsume lower-order ones. For example, convex sets may subsume intervals and points. The reverse is not true, however. This suggests that lower-order representations may be sacrificing robustness in favor of computational simplicity.

E. SUMMARY

The evidential-reasoning research reported in Chapter II can be summarized as follows:

- The evidential-reasoning problem can be formulated in terms of a four-part paradigm. The component parts are the background elements, the observation reports, the updating mechanism, and the decision mechanism.
- Each of the six major approaches can be expressed in terms of the four-part paradigm.
- Major similarities in the ER approaches are found in two background elements:
    - -- structure of the algebra of statements (but not necessarily the content)
    - -- the loss function.
- Major differences in the ER approaches are found in several components:
    - -- structures given to belief states (points, intervals, convex sets, fuzzy sets)
    - -- updating algorithms (Bayes' Theorem, Dempster's Rule, principles of direct inference, confidence intervals, fuzzy combination)
    - -- decision algorithms (expected loss on point-valued p-functions, expected loss on intervals or convex sets, fuzzy decision rules).

III: APPLICATION TO EXPERT-SYSTEM TASKS IN THE DOMAIN OF IMAGE ANALYSIS

A. INTRODUCTION

In Chapter II we formulated the relevant ER approaches in a parallel fashion and arrived at a set of points of correspondence and points of incommensurability. In Chapter III we apply the ER approaches to several important tasks for expert systems in the domain of image analysis.

We shall limit ourselves here to discussion of image analysis or low-level vision. The current effort also has application to scene analysis or high-level vision, but exploration of these applications lies beyond the scope of this effort. We shall consider the domain of image analysis to include the identification of lines, edges, regions, and texture through the level of the primal sketch (M2). We shall consider scene analysis to be the recognition of objects and configurations of objects based upon the results of image analysis.

Image analysis here will be considered to be comprised of three interrelated processes: feature extraction, segmentation, and domain classification. Feature extraction is the retrieval of relevant information from the original image data. Segmentation is the division of the image into several relevant domains. Domain classification is the assignment of each domain to one of several categories.

There are three expert-system tasks to be addressed here: diagnosis, integration, and control. The nature and importance of each will be described in terms of image-analysis objectives in the next section.

B. EXPERT-SYSTEM TASKS IN IMAGE ANALYSIS

There is growing evidence that the application of expert-system techniques to image-analysis problems may be of significant utility (for examples, see references B1, G3, K2, M1, M2, N2, and W1). Various algorithms for attacking various components of this problem have incorporated specialized heuristics to improve performance. The promise of additional improvement via expert systems lies in applying several classes of such heuristics in an integrated and adaptive manner to a single image.

While exploring the nature of the three major expert-system tasks in terms of image-analysis problems, this discussion is not intended to be exhaustive in terms of either expert-system or image-analysis techniques. Its intent is to serve as a vehicle for comparison the the evidential-reasoning approaches described in Chapter II.

1. Diagnosis

Diagnosis is the inference of system behaviour from various data. That is, based upon data about system processes, a description of the system state is constructed. Figure III-1 is a schematic representation of such a diagnostic task.

In terms of the component parts of the image-analysis process, diagnosis might include the following sorts of tasks:

- Assessment of the degree to which the image has been appropriately segmented
- Assessment of the degree to which domains have been appropriately classified
- Identification of those feature-extraction, segmentation, or domain-classification components of the image-analysis system that are not performing adequately in a certain context.

Detailed discussion of each of these tasks is beyond the scope of this effort, but we will construct an example for use in the remainder of the chapter. We will focus on the region-growing (RG) portion of segmentation.

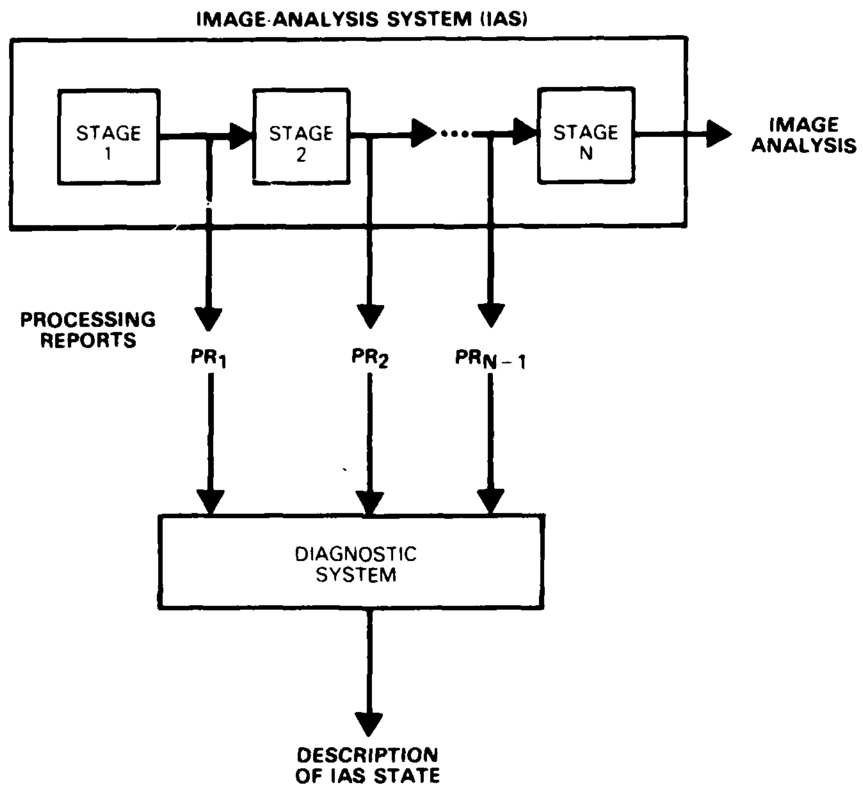


FIGURE III-1.
SCHEMATIC REPRESENTATION OF DIAGNOSTIC TASKS

Sample Task I: Diagnosis

Objective:	Diagnose the behavior of the RG component in the segmentation portion of the IAS.
Assumptions:	RG operates on features of intensity and intensity gradient.
Inputs:	Measures of the uniformity of each current region Measures of strengths of current boundaries
Outputs:	Statements concerning RG behavior, e.g., S1 - "The RG is merging too many pixels." S2 - "The RG is merging approximately the right number of pixels." S3 - "The RG is merging too few pixels."
Approach:	(1) Assemble uniformity and boundary strength measures. (2) Infer support for hypotheses corresponding to statements about RG behaviour. (3) Select a statement for output.

Additional inputs could be used in the diagnostic process if RG were to operate on features such as color hue, color saturation, or region size. Our current intent in use of the sample task is well served by limiting it to the simple inputs described.

2. Integration

Integration may be interpreted as the meaningful combination of a number of disparate inputs into a smaller number of outputs. For example, given assignments of a characteristic to some object obtained via distinct methods, combine the assignments to form one integrated assignment. Figure III-2 is a schematic representation of such an integrative task.

In terms of the image-analysis process, integration includes the following sorts of tasks:

- Combining two or more classification maps into a single map
- Combining two or more feature vectors into a single vector prior to classification.

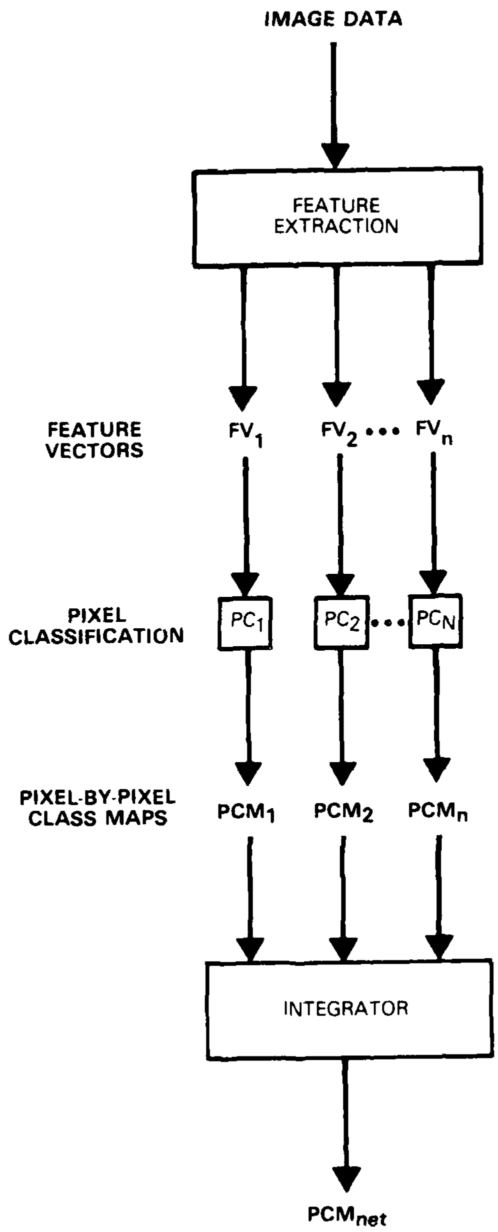


FIGURE III-2.
SCHEMATIC REPRESENTATION OF INTEGRATION TASKS

Detailed discussion of these tasks is again beyond our current scope. The example we will use in the remainder of the chapter will be concerned with the integration of the results of two different pixel-classification maps.

Sample Task II: Integration

Objective: Combine two pixel-classification maps into a single map.

Assumptions: Each map is classified into the same six classes, $(C_1, \ldots, C_6)$.

Classification strengths have one of four values, $(S_1, \ldots S_4)$.

Inputs: Two 1024 x 1024 classification maps in which each pixel has been assigned to one of six classes. Each assignment carries a strength.

Outputs: One 1024 x 1024 classification map in which each pixel has been assigned to one of six classes. Each assignment carries a strength.

Approach:
(1) Assemble classification maps.
(2) Combine the following functions:

$$M_1(x,y) = (C_i, S_j)$$
$$M_2(x,y) = (C_i, S_j) \quad ,$$

into a third function of the same form, M_{12}.

(3) Present M_{12} as the integrated result.

As before, this task is purposely kept simple for clarity of analysis.

3. Control

Control may be interpreted as the choice of actions that influence system behavior. That is, based upon a description of system state, a decision process is followed that results in recommended actions designed to bring system behavior closer to some goal state. Figure III-3 is a schematic representation of control tasks.

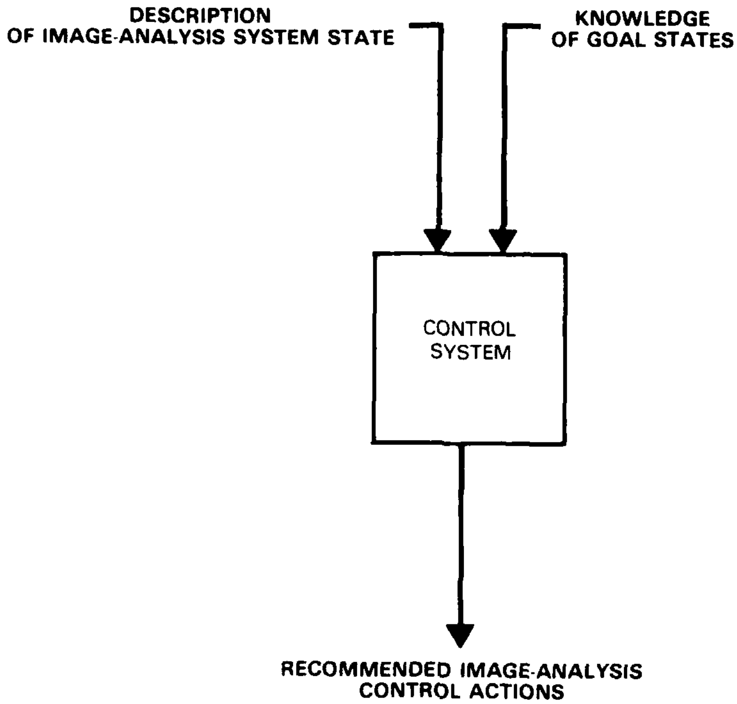


FIGURE III-3.
SCHEMATIC REPRESENTATION OF CONTROL TASKS

509-2-14-85-1

In the image-analysis process, control includes the following sorts of tasks:

- Determination of the appropriate features to be extracted based upon the nature of the image data, the performance of components of the image-analysis system, and other relevant factors.
- Choice of methods of component reduction (ETL-0343, p.7).
- Choice of raster-processing techniques (ETL-0347, p.7).
- Choice of interest operators (ETL-0347, p.20).
- Choice of weights and compatibility values for relaxation-labeling techniques (ETL-0280, p.9).
- Removal of artifacts at class boundaries (ETL-0300, p.19).
- Coordination of cooperative algorithms (ETL-0305, p.5).

- Choice of rules for coordination of concurrent algorithms (ETL-0298, p.20).

Of course, detailed discussion of these tasks is beyond our current scope. We will construct an example for use in the remainder of the chapter that is concerned with the coordination of two algorithms, one dealing with lines and the other with regions.

Sample Task III: Control

Objective: Coordinate the application of two cooperative image-analysis algorithms, line removal (LR) and region splitting (RS)

Assumptions: LR and RS interact only via two rules:

R1 - IF region y is not small
& region y is bisected by line x
& line x is not short
& the average gradient of line x is high

THEN split region y along line x.

R2 - IF line x is incomplete
& the average gradient of line x is low
& region y lies on both sides of line x,
THEN remove line x.

The location of the system in state-space is imprecisely known due to local, rather than global, measurements on regions and lines.

Inputs: Measures on region y - size.
Measures on line x - completeness, length, average gradient.
Measures between region y and line x - bisection.

Outputs: Recommended actions:
A1 - Apply R1 to region y.
A2 - Do not apply R1 to region y.
A3 - Apply R2 to line x.
A4 - Do not apply R2 to line x.

Approach: (1) Assemble measures.
(2) Infer support for antecedents of R1 and R2.
(3) Select recommended action.

It should be noted that we have constructed rules R1 and R2 so that the antecedents cannot be simultaneously satisfied for both rules. This simplifies the control task in that no rules to choose between R1 and R2 are required. Rules about the use of other rules are often termed meta-rules. Meta-rules can be constructed in much the same format as R1 and R2; the application of the evidential-reasoning approaches would proceed in essentially the same fashion for such rules.

C. APPLICATION OF EVIDENTIAL-REASONING APPROACHES

Having identified three major categories of expert-system tasks and constructed samples of such tasks, we are ready to explore the application of the evidential-reasoning approaches to these sample tasks. In this section, we outline application of each approach in turn, taking care to preserve parallel, step-wise treatments. We postpone discussion until Section D, where we compare the results. These outlines are designed to be precursors of the pseudo-code required to perform computational experiments in subsequent work. It is clear, however, that they are here presented in much simplified form in order to cover the great breadth of the current effort (6 approaches and 3 tasks).

1. Approach 1: Classical Bayes

a. Sample Task I - Diagnosis

Application of the classical Bayes approach would proceed as follows:

(1) Determine elements of the algebra of statements:
- values of intensity for each relevant pixel
- values of intensity gradient for each relevant pixel
- set of current regions
- measure of uniformity for each current region
- measures of boundary strengths.

(2) Determine conditional p-values (point-valued).

(3) Determine a priori p-values (point-valued).

(4) Identify those statements having unitary p-values on the basis of the input data.

(5) Apply equations (1.1) and (1.2) to obtain a posteriori and current p-values.

(6) Determine loss matrix.

(7) Calculate the expected loss using equation (1.3) based upon the current p-values.

(8) Select the diagnosis statement having the lowest expected loss.

b. <u>Sample Task II - Integration</u>

Application of the classical Bayes approach would proceed as follows:

(1) Determine elements of the algebra of statements:
  - 24 states of a pixel from Map 1, (C_1, S_1)
  - 24 states of a pixel from Map 2, (C_2, S_2)
  - 24 states of a pixel from Map 12, (C_{12}, S_{12})

(2) Determine conditional p-values:
  - point-valued
  - values for $P[(C_1,S_1)\ (C_2,S_2)\ |\ (C_{12}, S_{12})]$
  - up to 24^3 such values, but presumably many are zero.

(3) Determine a priori p-values.
  - point-valued
  - 24 values for $P_{prior}[(C_{12}, S_{12})]$

(4) Identify those statements having unitary p-values on the basis of the input data.
  - If each pixel is considered independently, just two statements will have unitary p-values.

(5) Apply equations (1.1) and (1.2) to obtain a posteriori and current p-values.
  - 24 values for $P_{post}[(C_{12}, S_{12})]$.

(6) Determine loss matrix.

(7) Calculate the expected loss using equation (1.3) based upon the current p-values.

(8) Select the statement for the Map-12 pixel having the lower expected loss.

c. Sample Task III - Control

Application of the classical Bayes approach would proceed as follows:

(1) Determine elements of the algebra of statements.

- Relative to the observation space:
  - -- r states of size measure on region y
  - -- s states of completeness measure on line x
  - -- t states of length measure on line x
  - -- u states of average-gradient measure on line x
  - -- v states of bisection measure on region y and line x
  - -- there are thus N states in OS, the observation space. N is the product of r, s, t, u, and v.
- Relative to the control space:
  - -- 2 states for each of the five measures above
  - -- There are thus 32 states in CS, the control space.

(2) Determine conditional p-values:

- point-valued
- values for $P[OS_j \mid CS_i]$
- up to (32 x N) such values, but presumably many are zero.

(3) Determine a priori p-values.

- point-valued
- 32 values for $P_{prior}[CS_i]$.

(4) Identify those statements having unitary p-values on the basis of the input data:

- For a given region/line pair, just one statement in OS will have a unitary p-value.

(5) Apply equations (1.1) and (1.2) to obtain a posteriori and current p-values.

- 32 values for $P_{post}[CS_i]$

(6) Determine loss matrix.

(7) Calculate the expected loss using equation (1.3) based upon the current p-values.

(8) Select the control action having the lower expected loss.

2. Approach 2: Convex Bayes

a. Sample Task I - Diagnosis

Application of the convex Bayes approach would proceed as follows:

(1) Determine elements of the algebra of statements.

- values of intensity for each relevant pixel
- values of intensity gradient for each relevant pixel
- set of current regions
- measure of uniformity for each current region
- measures of boundary strengths.

(2) Determine conditional p-values:

- intervals or convex sets.

(3) Determine a priori p-values:

- intervals or convex sets.

(4) Identify those statements having non-null p-values on the basis of the input data.

(5) Apply Bayes' Theorem to obtain a posteriori p-values:

- must use interval or set form of theorem, equation (2.1).

(6) Determine loss matrix.

(7) Select the diagnosis statement via some decision function that uses the loss matrix and the current p-values.

b. Sample Task II - Integration

Application of the convex Bayes approach would proceed as follows:

(1) Determine elements of the algebra of statements:

- 24 states of a pixel from Map 1, (C_1, S_1)
- 24 states of a pixel from Map 2, (C_2, S_2)
- 24 states of a pixel from Map 12, (C_{12}, S_{12}).

(2) Determine conditional p-values.

- intervals or convex sets
- values for $P[(C_1,S_1)\ (C_2,S_2) \mid (C_{12}, S_{12})]$
- there could be up to 24^3 such intervals or convex sets, but presumably many are null.

(3) Determine a priori p-values:

- intervals or convex sets
- 24 intervals or sets for $P_{prior}[(C_{12}, S_{12})]$

(4) Identify those statements having non-null p-values on the basis of the input data.

- if each pixel is considered independently, just two statements will have non-null p-values.

(5) Apply Bayes' Theorem to obtain a posteriori p-values:

- 24 intervals or sets for $P_{post}[(C_{12}, S_{12})]$
- must use interval or set form of theorem, equation (2.1).

(6) Determine loss matrix.

(7) Select the statement for the Map-12 pixel via some decision function that uses the loss matrix and the current p-values.

c. Sample Task III - Control

Application of the convex Bayes approach would proceed as follows:

(1) Determine elements of the algebra of statements:

- Relative to the observation space:
  - -- r states of size measure on region y
  - -- s states of completeness measure on line x
  - -- t states of length measure on line x
  - -- u states of average-gradient measure on line x
  - -- v states of bisection measure on region y and line x
  - -- there are thus N states in OS, the observation space. N is the product of r, s, t, u, and v.
- Relative to the control space:
  - -- 2 states for each of the five measures above
  - -- there are thus 32 states in CS, the control space.

(2) Determine conditional p-values:

- intervals or convex sets
- values for $P[OS_j \mid CS_i]$
- there could be up to (32 x N) such intervals or convex sets, but presumably many are null.

(3) Determine a priori p-values:

- intervals or convex sets
- 32 intervals or sets for $P_{prior}[CS_i]$.

(4) Identify those statements having non-null p-values on the basis of the input data:

- for a given region/line pair, just one statement in OS will have a non-null p-value.

(5) Apply Bayes' Theorem to obtain a posteriori p-values.

- 32 intervals or sets for $P_{post}[CS_i]$.
- must use interval or set form of theorem, equation (2.1).

(6) Determine loss matrix.

(7) Select the control action via some decision function that uses the loss matrix and the current p-values.

3. Approach 3: Dempster - Shafer

a. Sample Task I - Diagnosis

Application of the Dempster-Shafer approach would proceed as follows:

(1) Determine elements of the algebra of statements:

- values of intensity for each relevant pixel
- values of intensity gradient for each relevant pixel
- set of current regions
- measure of uniformity for each current region
- measures of boundary strengths

(2) Determine mass functions for statements on the basis of the input data.

(3) Apply Dempster's Rule using equation (3.4) if combination of mass functions is required.

(4) Compute support and plausibility via equations (3.1) and (3.2) for hypotheses corresponding to the diagnosis statements.

(5) Determine loss matrix.

(6) Select the diagnosis statement via some decision function that uses the loss matrix, the support function, and the plausibility function.

b. Sample Task II - Integration

Application of the Dempster-Shafer approach would proceed as follows:

(1) Determine elements of the algebra of statements.

- 24 states of a pixel from Map 1, (C_1, S_1)
- 24 states of a pixel from Map 2, (C_2, S_2)
- 24 states of a pixel from Map 12, (C_{12}, S_{12})

(2) Determine mass functions for statements on the basis of the input data:

- if each pixel is considered independently, just two statements will have non-null mass-functions.

(3) Apply Dempster's Rule using equation (3.4) if combination of mass functions is required:

- combine the two non-null mass functions.

(4) Compute support and plausibility via equations (3.1) and (3.2) for statements corresponding to the states of the Map-12 pixel.

(5) Determine loss matrix.

(6) Select the statement for the Map-12 pixel via some decision function that uses the loss matrix, the support function, and the plausibility function.

c. Sample Task III - Control

Application of the Dempster-Shafer approach would proceed as follows:

(1) Determine elements of the algebra of statements:

- Relative to the observation space:
  - -- r states of size measure on region y
  - -- s states of completeness measure on line x
  - -- t states of length measure on line x
  - -- u states of average-gradient measure on line x

-- v states of bisection measure on region y and line x

-- there are thus N states in OS, the observation space. N is the product of r, s, t, u, and v.

- Relative to the control space:

-- 2 states for each of the five measures above

-- there are thus 32 states in CS, the control space.

(2) Determine mass functions for statements on the basis of the input data:

- for a given region/line pair, more than one statement in OS may have a non-null mass-function.

(3) Apply Dempster's Rule using equation (3.4) if combination of mass functions is required:

- combine the non-null mass functions.

(4) Compute support and plausibility via equations (3.1) and (3.2) for statements corresponding to the control actions.

(5) Determine loss matrix.

(6) Select statement for the control action via some decision function that uses the loss matrix, the support function, and the plausibility function.

4. Approach 4: Kyburg

a. Sample Task I - Diagnosis

Application of the Kyburg approach would proceed as follows:

(1) Determine elements of the algebra of statements.

- values of intensity for each relevant pixel
- values of intensity gradient for each relevant pixel
- set of current regions
- measure of uniformity for each current region
- measures of boundary strengths

(2) Apply principle of direct inference.

- K-relevance

(3) Determine conditional p-values.

- intervals

(4) Determine a priori p-values.

- intervals

(5) Identify those statements having non-null p-values on the basis of the input data.

(6) Apply direct inference to obtain a posteriori p-values.

(7) Determine loss matrix.

(8) Select the diagnosis statement via some decision function that uses the loss matrix and the a posteriori p-values.

b. Sample Task II - Integration

Application of the Kyburg approach would proceed as follows:

(1) Determine elements of the algebra of statements.

- 24 states of a pixel from Map 1, (C_1, S_1)
- 24 states of a pixel from Map 2, (C_2, S_2)
- 24 states of a pixel from Map 12, (C_{12}, S_{12}).

(2) Apply principle of direct inference.

- K-relevance

(3) Determine conditional p-values.

- intervals
- values for $P[(C_1,S_1)\ (C_2,S_2) \mid (C_{12}, S_{12})]$
- there could be up to 24^3 such intervals, but presumably many are null.

(4) Determine a priori p-values:

- intervals
- 24 intervals for $P_{prior}[(C_{12}, S_{12})]$.

(5) Identify those statements having non-null p-values on the basis of the input data:

- if each pixel is considered independently, just two statements will have non-null p-values.

(6) Apply direct inference to obtain a posteriori p-values:

- 24 intervals for $P_{post}[(C_{12}, S_{12})]$
- must use interval form of theorem.

(7) Determine loss matrix.

(8) Select the statement for the Map-12 pixel via some decision function that uses the loss matrix and the a posteriori p-values.

c. Sample Task III - Control

Application of the Kyburg approach would proceed as follows:

(1) Determine elements of the algebra of statements:

- Relative to the observation space:
  - -- r states of size measure on region y
  - -- s states of completeness measure on line x
  - -- t states of length measure on line x
  - -- u states of average-gradient measure on line x
  - -- v states of bisection measure on region y and line x
  - -- there are thus N states in OS, the observation space. N is the product of r, s, t, u, and v.
- Relative to the control space:
  - -- 2 states for each of the five measures above
  - -- there are thus 32 states in CS, the control space.

(2) Apply principle of direct inference:

- K-relevance.

(3) Determine conditional p-values:

- intervals
- values for $P[OS_j \mid CS_i]$
- There could be up to (32 x N) such intervals, but presumably many are null.

(4) Determine a priori p-values.

- intervals
- 32 intervals for $P_{prior}[CS_i]$.

(5) Identify those statements having non-null p-values on the basis of the input data.

- for a given region/line pair, just one statement in OS will have a non-null p-value.

(6) Apply direct inference to obtain a posteriori p-values.

- 32 intervals for $P_{post}[CS_i]$
- must use interval form of theorem.

(6) Determine loss matrix.

(7) Select the control action via some decision function that uses the loss matrix and the a posteriori p-values.

5. <u>Approach 5: Neyman-Pearson</u>

a. <u>Sample Task I - Diagnosis</u>

Application of the Neyman-Pearson approach would proceed as follows:

(1) Determine elements of the algebra of statements:

- values of intensity for each relevant pixel
- values of intensity gradient for each relevant pixel
- set of current regions
- measure of uniformity for each current region
- measures of boundary strengths

(2) Choose confidence level, L.

(3) Construct confidence sets for the *diagnosis statements:*

- interval-valued

(4) Identify those observation statements having unitary p-values on the basis of the input data.

(5) Determine the current limits on the p-values of the diagnosis statements via the confidence sets pertaining to the observation statements identified in (4).

(6) Determine loss matrix.

(7) Select the diagnosis statement via a decision function that uses the limits generated in (5).

b. <u>Sample Task II - Integration</u>

Application of the Neyman-Pearson approach would proceed as follows:

(1) Determine elements of the algebra of statements.

- 24 states of a pixel from Map 1, (C_1, S_1)
- 24 states of a pixel from Map 2, (C_2, S_2)
- 24 states of a pixel from Map 12, (C_{12}, S_{12})

(2) Choose confidence level, L.

(3) Construct confidence sets for the Map-12 statements.

- interval-valued.

(4) Identify those observation statements having unitary p-values on the basis of the input data.

(5) Determine the current limits on the p-values of the Map-12 statements via the confidence sets pertaining to the observation statements identified in (4).

(6) Determine loss matrix.

(7) Select the Map-12 statement via a decision function that uses the limits generated in (5).

c. <u>Sample Task III - Control</u>

Application of the Neyman-Pearson approach would proceed as follows:

(1) Determine elements of the algebra of statements:

- Relative to the observation space:
  - -- r states of size measure on region y
  - -- s states of completeness measure on line x
  - -- t states of length measure on line x
  - -- u states of average-gradient measure on line x
  - -- v states of bisection measure on region y and line x
  - -- There are thus N states in OS, the observation space. N is the product of r, s, t, u, and v.
- Relative to the control space:
  - -- 2 states for each of the five measures above
  - -- there are thus 32 states in CS, the control space.

(2) Choose confidence level, L.

(3) Construct confidence sets for the control-space statements.

- interval-valued

(4) Identify those observation-space statements having unitary p-values on the basis of the input data.

(5) Determine the current limits on the p-values of the control-space statements via the confidence sets pertaining to the observation-space statements identified in (4).

(6) Determine loss matrix.

(7) Select the control-space statement via a decision function that uses the limits generated in (5).

6. Approach 6: Possibility

a. Sample Task I - Diagnosis

Application of the possibility approach would proceed as follows:

(1) Determine elements of the algebra of statements.

- values of intensity for each relevant pixel
- values of intensity gradient for each relevant pixel
- set of current regions
- measure of uniformity for each current region
- measures of boundary strengths

(2) Construct membership functions.

- two-element fuzzy sets linking parameters above to each diagnosis statement

(3) Set values in the membership functions on the basis of the input data.

(4) Combine the membership functions via one or more of the fuzzy decision functions, equations (6.3), (6.4), (6.5).

(5) Determine goal and constaint fuzzy sets via equations (6.6) and (6.7).

(6) Form the confluence set, DEC(G,C) via equation (6.8).

(7) Select the diagnosis statement having the highest DEC degree of membership.

b. Sample Task II - Integration

Application of the possibility approach would proceed as follows:

(1) Determine elements of the algebra of statements.

- 24 states of a pixel from Map 1, (C1, S1)
- 24 states of a pixel from Map 2, (C2, S2)
- 24 states of a pixel from Map 12, (C12, S12)

(2) Construct membership functions.

- 24 two-element fuzzy sets linking Map-1 and Map-2 parameters to each Map-12 statement

(3) Set values in the membership functions on the basis of the input data.

(4) Combine the membership functions via one or more of the fuzzy decision functions, equations (6.3), (6.4), (6.5).

(5) Determine goal and constaint fuzzy sets via equations (6.6) and (6.7).

(6) Form the confluence set, DEC(G,C) via equation (6.8).

(7) Select the Map-12 statement having the highest DEC degree of membership.

c. Sample Task III - Control

Application of the possibility approach would proceed as follows:

(1) Determine elements of the algebra of statements:

- Relative to the observation space:
  - -- r states of size measure on region y
  - -- s states of completeness measure on line x
  - -- t states of length measure on line x
  - -- u states of average-gradient measure on line x
  - -- v states of bisection measure on region y and line x
  - -- There are thus N states in OS, the observation space. N is the product of r, s, t, u, and v.
- Relative to the control space:
  - -- 2 states for each of the five measures above
  - -- There are thus 32 states in CS, the control space.

(2) Construct membership functions:

- 32 N-element fuzzy sets linking observation-space parameters to each control-space state

(3) Set values in the membership functions on the basis of the input data.

(4) Combine the membership functions via one or more of the fuzzy decision functions, equations (6.3), (6.4), (6.5).

(5) Determine goal and constaint fuzzy sets via equations (6.6) and (6.7).

(6) Form the confluence set, DEC(G,C) via equation (6.8).

(7) Select the control action having the highest DEC degree of membership.

D. COMPARATIVE ANALYSIS OF THE APPLICATIONS

From the foregoing definition of sample tasks and parallel, step-wise application of theoretical approaches, we can make the following comparative observations:

- The diagnosis and integration tasks do not lend themselves to formulation of loss or utility functions as readily as the control task. This occurs because the former tasks, at least when considered in isolation, do not ordinarily incorporate consequences of decisions in their formulation. In a sense, they do not make explicit the impact of Type-I and Type-II errors. This difference in tasks may favor approaches that do not explicity require a loss function. However, in practical applications where the diagnosis and integration tasks are embedded in a specific IAS, it may be the case that useful loss functions can be constructed by reference to the role of the task in the overall performance of the system.

- The control task offers the broadest spectrum of possible applications in image-analysis. This comes about because the process of control requires the broadest view of system goals, procedures, and assumptions. Control can ordinarily subsume both diagnosis and integration. It will, however, require a large algebra of statements and will thus penalize those approaches that extend this algebra to include the broader representations of belief states.

- The classical Bayes approach offers a somewhat simpler and more familiar technique of evidential calculation. As was seen in the sample tasks, appropriate conditional and a priori p-values must be developed. Furthermore, their influence permeates the entire ER process; if they are inapproapriate, then the process may not yield the desired result.

- The convex Bayes approach offers a more robust representation of belief states capable of handling each of the sample tasks. The additional computation required may be justified when addressing complex tasks such as control, but is probably not justified when addressing simpler tasks such as integration. In addition, the overhead associated with sorting out the proper use of expected loss in this context may only be palatable when the task is sufficiently complex.

- The Dempster-Shafer approach offers a somewhat more robust belief-state representation and also appears to be capable of handling each of the sample tasks. Comments concerning justification of the additional computational burden also apply here, although some research seeking efficient algorithms has appeared (B2). The proper translation of support and plausibility into decisions is not yet based upon a large body of practical experience, and thus adds some difficulty to possible applications in image analysis. In addition, the influence of the initial mass distribution on the evolution of belief states requires detailed investigation.

- The Kyburg approach is the most abstract, but may offer significant benefits when tasks involve sorting relevant from irrelevant knowledge. The K-relevance criterion has had little exercise in the world of practical applications, but studies such as this one form a growing basis for such application.

- The Neyman-Pearson approach offers a concept, the confidence interval, that is initially appealing in terms of modeling the confidence that human experts might have in the evidential connection between various parameters. However, its foundations are clearly limited to the treatment of statistical hypotheses involving well-defined populations. This fact advises caution in the extension of this approach to less statistically-oriented domains. Within domains that correspond to sampling tasks, its application may well be justified.

- The possibility approach offers a framework within which it may be more convenient to capture certain relevant knowledge in the image-analysis domain, particularly in complex control tasks. This is based upon its ability to characterize imprecise linguistic terms ("usually", "sometimes", etc.) as fuzzy sets. Initial results are promising (M2), but require extension.

Directions for further research based upon these observations will be suggested in Chapter IV.

E. SUMMARY

The research into application of evidential-reasoning approaches to expert-system tasks in image analysis reported in Chapter III can be summarized as follows:

- Major expert-system tasks in this domain are: (1) diagnosis, the inference of system behavior from data on system processes, (2) integration, the meaningful combination of a number of disparate inputs into a smaller number of outputs, and (3) control, the choice of actions that influence system behavior.

- Each of the ER approaches can be applied to sample tasks from these three categories. Several strengths and weaknesses can be identified:

  - Interval and convex-set representations of belief states may be useful in complex ES tasks (e.g., control), but do so at the expense of added complexity.

  - New and more general decision procedures must be developed in order to make practical use of these robust representations.

  - Criteria of evidential relevance are being developed, but require practical application for assessment.

  - Imprecise linguistic terms may be characterized by fuzzy sets, but this also requires practical application for assessment.

Chapter IV summarizes overall results from the current effort and discusses directions for futher research.

IV: CURRENT RESULTS AND DIRECTIONS FOR FURTHER RESEARCH

A. CURRENT RESULTS

The evidential-reasoning research reported in Chapter II can be summarized as follows:

- The evidential-reasoning problem can be formulated in terms of a four-part paradigm. The component parts are the background elements, the observation reports, the updating mechanism, and the decision mechanism.
- Each of the six major approaches can be expressed in terms of the four-part paradigm.
- Major similarities in the ER approaches are found in two background elements:
  - -- structure of the algebra of statements (but not necessarily the content)
  - -- the loss function.
- Major differences in the ER approaches are found in several components:
  - -- structures given to belief states (points, intervals, convex sets, fuzzy sets)
  - -- updating algorithms (Bayes' Theorem, Dempster's Rule, principles of direct inference, confidence intervals, fuzzy combination)
  - -- decision algorithms (expected loss on point-valued p-functions, expected loss on intervals or convex sets, fuzzy decision rules).

The research into application of evidential-reasoning approaches to expert-system tasks in image-analysis reported in Chapter III can be summarized as follows:

- Major expert-system tasks in this domain are: (1) diagnosis, the inference of system behavior from data on system processes, (2) integration, the meaningful combination of a number of disparate inputs into a smaller number of outputs, and (3) control, the choice of actions that influence system behavior.

- Each of the ER approaches can be applied to sample tasks from these three categories. Several strengths and weaknesses can be identified:

  -- Interval and convex-set representations of belief states may be useful in complex ES tasks (e.g., control), but do so at the expense of added complexity.

  -- Specialized decision procedures must be developed in order to make practical use of these robust representations.

  -- Criteria of evidential relevance are being developed, but require practical application for assessment.

  -- Imprecise linguistic terms may be characterized by fuzzy sets, but this also requires practical application for assessment.

In the following section we suggest additional research based upon these results.

B. DIRECTIONS FOR FURTHER RESEARCH

The current effort has resulted in identification of several key issues for the application of expert systems to image analysis. The focal points of these issues are:

- Nature of the most important IAS tasks - precise delineation of the task objectives, scope, assumptions, inputs, outputs, and approach is required.
- Utility of interval or convex-set representation of belief states
- Means of representation of relevant knowledge using the rule-oriented paradigm.

It is clear that these issues are interrelated. However, at this stage of development of expert systems for image analysis, it appears that the most important is the first; without such definition of real tasks, the research issues remain too broad to attack efficiently.

Presented below are outlines of representative approaches to further research and development. These are designed to embody the lessons learned thus far and to extend efforts in a practical manner:

- R&D Plan I

  -- Objective: develop a prototype image-analysis expert system for application to current ETL concerns.

  -- Approach:
     (1) identify specific image-analysis tasks.
     (2) detail inputs, outputs, goals, etc.
     (3) identify and characterize available hardware and software environments.
     (4) identify relevant knowledge.
     (5) structure knowledge-base using rule-oriented, object-oriented, procedure-oriented, and data-oriented paradigms, as appropriate.
     (6) develop inference control system that will utilize the knowledge-base.
     (7) exercise prototype on sample tasks.
     (8) revise knowledge-base and inference control system, as appropriate.
     (9) test and evaluate prototype on real tasks.

Plan I assumes that further research into the choice of appropriate eviden-tial-resoning techniques is embedded in step (6) of the approach. Plan II concentrates on this research:

- R&D Plan II

  -- Objective: Compare several ER techniques in direct application to one or more detailed image-analysis tasks.

  -- Approach:
     (1) identify specific image-analysis tasks.
     (2) detail inputs, outputs, goals, etc.
     (3) identify rules that govern appropriate performance of the image-analysis tasks.
     (4) choose subset of the ER approaches for detailed numeric application.
     (5) perform sample calculations using data supplied by ETL (or in a format specified by ETL).

(6) compare results in terms of desired objectives, computational burden, and sensitivity to variations in data and assumptions.

These two plans are representative of the research directions available, but not exhaustive of them. Other plans might: (1) concentrate on developing rules for specific image-analysis tasks, (2) investigate the relative utility of various types of rule-based control systems, or (3) investigate the utility of trainable or learning expert-systems for image analysis.

GLOSSARY

ER	Evidential Reasoning
ES	Expert System
IAS	Image Analysis System
LR	Line Removal
LT	Lindenbaum-Tarski Algebra
PLS	Plausibility
PR	Pattern Recognition
RG	Region Growth
RS	Region Split
SPT	Support

REFERENCES

B1 Ballard, D., Brown, C., and Feldman, J. "An Approach to Knowledge-Directed image Analysis." In Proceedings of the Fifth International Joint Conference on Artificial Intelligence, pp. 664-670, 1977.

B2 Barnett, J. "Computational Methods for a Mathematical Theory of Evidence." In Proceedings of the Seventh International Joint Conference on Artificial Intelligence, pp. 868-875, 1981.

B3 Bayes, T. "An Essay Toward Solving a Problem in the Doctrine of Chances." Phil. Trans. Royal Soc., 330-418, (1763). Reprinted in Biometrika, 45, 293-315, (1958).

B4 Bellman, R., and Zadeh, L. Decision-Making in a Fuzzy Environment. Technical Report ERL-69-8, Electronics Research Laboratory, University of California, Berkeley, 1969.

B5 Blackwell, D., and Girshick, M. Theory of Games and Statistical Decisions, Wiley, New York, 1954.

C1 Castleman, K. Digital Image Processing, Prentice-Hall, Englewood Cliffs, New Jersey, 1979.

C2 Chernoff, H., and Moses, L. Elementary Decision Theory, Wiley, New York, 1959.

C3 Cohen, P., and Feigenbaum, E., eds. The Handbook of Artificial Intelligence: Volume 3, Kaufmann, Los Altos, Calif., 1982.

C4 Crombie, M., and Shine, J. An Analysis of a Relaxation Scheme to Improve Terrain Elevation Data. Report ETL-0298, U.S. Army Corps of Engineers, Engineer Topographic Laboratories, Fort Belvoir, Virginia, 1982.

C5 Crombie, M., Rand, R., and Friend, N. An Analysis of the Max-Min Texture Measure. Report ETL-0280, U.S. Army Corps of Engineers, Engineer Topographic Laboratories, Fort Belvoir, Virginia, 1982.

C6 ------ Scene Classification Results Using the Max-Min Texture Measure. Report ETL-0300, U.S. Army Corps of Engineers, Engineer Topographic Laboratories, Fort Belvoir, Virginia, 1982.

C7 ------ Feature Component Reduction Through Divergence Analysis. Report ETL-0305, U.S. Army Corps of Engineers, Engineer Topographic Laboratories, Fort Belvoir, Virginia, 1982.

D1 Dempster, A.P. "Upper and Lower Probabilities Induced by a Multivalued Mapping." Ann. Math. Stat., 38, 325-39, (1967).

D2 Dempster, A.P. "A Generalization of Bayesian Inference." J. Royal Stat. Soc., 30(2), 205-247, (1968).

D3 Dillard, R.A. Computing Probability Masses in Rule-Based Systems. Technical Document 545, Naval Ocean Systems Center, San Diego, 1982.

D4 Duda, R., and Hart, P. Pattern Classification and Scene Analysis, Wiley, New York, 1973.

D5 Duda, R., Hart, P., and Nilsson, N. "Subjective Bayesian Methods for Rule-Based Inference Systems." In Proceedings of AFIPS 1976 National Computer Conference, pp. 1075-1082, 1976.

F1 Friend, N. Analysis of Interactive Image Cleansing via Raster-Processing Techniques. Report ETL-0347, U.S. Army Corps of Engineers, Engineer Topographic Laboratories, Fort Belvoir, Virginia, 1983.

F2 Faugeras, O., and Price, K. "Semantic Description of Aerial Images Using Stochastic Labeling." In Proceedings of the Fifth International Conference on Pattern Recognition, pp. 352-357, 1980.

F3 Fischler, M., et al. "Modeling and Using Physical Constraints in Scene Analysis." In Proceedings of the National Conference on Artificial Intelligence, pp. 30-35, 1982.

G1 Gaines, B. "Fundamentals of Decision: Probabilistic, Possibilistic, and other Forms of Uncertainty in Decision Analysis." In Fuzzy Sets and Decision Analysis, H. Zimmerman, L. Zadeh, and B. Gaines, eds., North-Holland, Amsterdam, pp. 47-65, 1984.

G2 Garvey, T., Lowrance, J., and Fischler, M. "An Inference Technique for Integrating Knowledge from Disparate Sources." In Proceedings of the Seventh International Joint Conference on Artificial Intelligence, pp. 319-325, 1981.

G3 Glicksman, J. "Using Multiple Information Sources in a Computational Vision System." In Proceedings of the Eighth International Joint Conference on Artificial Intelligence, pp. 1078-1080, 1983.

G4 Gupta, M., and Sanchez, E., eds. Approximate Reasoning in Decision Analysis, North-Holland, Amsterdam, 1982.

H1 Hayes-Roth, F. "Knowledge Representation, Organization, and Control in Large-Scale Pattern-Based Understanding Systems." IEEE Joint Workshop on Pattern Recognition and Artifical Intelligence, pp. 66-73, 1976.

H2 Hayes-Roth, F., Waterman, D., and Lenat, D. Building Expert Systems, Addison-Wesley, Reading, Mass., 1983.

H3 Huertas, A., and Nevatia, R. "Detection of Buildings in Aerial Images Using Shape and Shadows." In Proceedings of the Eighth International Joint Conference on Artificial Intelligence, pp. 1099-1101, 1983.

J1 Jeffrey, R. The Logic of Decision, 2nd edition, University of Chicago Press, Chicago, 1983.

J2 Jeffreys, H. Theory of Probability, 3rd edition, Oxford University Press, Oxford, 1961.

K1 Kendall, M., and Stuart, A. The Advanced Theory of Statistics: Volume 2--Inference and Relationship, 4th edition, Macmillan, New York, 1979.

K2 Kestner, W., et al. "Object-Guided Segmentation of Aerial Images." In Proceedings of the Fifth International Conference on Pattern Recognition, pp. 529-531, 1980.

K3 Kyburg, H.E., Jr. The Logical Foundations of Statistical Inference, Reidel, Boston, 1974.

K4 ------ Bayesian and Non-Bayesian Evidential Updating. Technical Report 139, University of Rochester Computer Science Department, 1984.

K5 Kullback, S. Information Theory and Statistics, Wiley, New York, 1959.

L1 Lehmann, E. Testing Statistical Hypotheses, Wiley, New York, 1959.

L2 Levi, I. The Enterprise of Knowledge, MIT Press, Cambridge, Mass., 1980.

L3 Lindley, D. Bayesian Statistics, A Review, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1972.

M1 Mackworth, A., and Havens, W. "Structuring Domain Knowledge for Visual Perception." In Proceedings of the Seventh International Joint Conference on Artificial Intelligence, pp. 625-627, 1981.

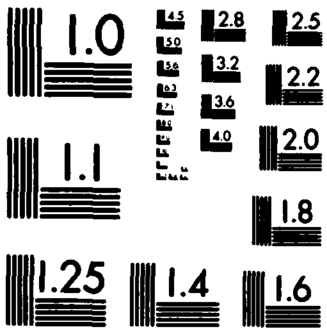
AD-A153 379 EVIDENTIAL REASONING IN EXPERT SYSTEMS FOR IMAGE ANALYSIS(U) SYSTEM PLANNING CORP ARLINGTON VA T R THOMPSON FEB 85 ETL-0381 DACA72-84-C-0006 2/2

UNCLASSIFIED F/G 9/2 NL

END

FILMED

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

M2 Mamdani, E., Ostergaard, J., and Lembessis, E. "Use of Fuzzy Logic for Implementing Rule-Based Control of Industrial Processes." In Fuzzy Sets and Decision Analysis, H. Zimmerman, L. Zadeh, and B. Gaines, eds., North-Holland, Amsterdam, pp. 429-445, 1984.

M3 Marr, D. Vision, Freeman, San Francisco, 1982.

M4 Medioni, G. "Segmentation of Images into Regions Using Edge Information." In Proceedings of the National Conference on Artificial Intelligence, pp. 42-45, 1982.

N1 Nagao, M., et al. "Structural Analysis of Complex Aerial Photographs." In Proceedings of the Sixth International Joint Conference on Artificial Intelligence, pp. 610-616, 1979.

N2 Nazif, A., and Levine, M. "Low Level Image Segmentation: An Expert System." IEEE Trans. on Pattern Analysis and Machine Intelligence, 6(5), 555-577, 1984.

N3 Neyman, J. "Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability." Phil. Trans. A, 236, 333 (1937).

R1 Rand, R., and Shine, J. Feature Analysis and Laws Texture Measure. Report ETL-0343, U.S. Army Corps of Engineers, Engineer Topographic Laboratories, Fort Belvoir, Virginia, 1983.

R2 Rosenfeld, A., and Kak, A. Digital Picture Processing, Volumes 1 & 2, 2nd edition, Academic, Orlando, 1982.

S1 Savage, L.J. The Foundations of Statistics, 2nd revised edition, Dover, New York, 1972.

S2 ------, et al. The Foundations of Statistical Inference--A Discussion, Methuen, London, 1962.

S3 Selfridge, P., and Sloan, K. "Reasoning About Images: Using Meta-Knowledge in Aerial Image Understanding." In Proceedings of the Seventh International Joint Conference on Artificial Intelligence, pp. 755-757, 1981.

S4 Shafer, G. A Mathematical Theory of Evidence, Princeton University Press, Princeton, N.J., 1976.

S5 Shneier, M. "Recognition Using Semantic Constraints." In Proceedings of the Fifth International Joint Conference on Artificial Intelligence, pp. 585-589, 1977.

S6 Strat, T. "Continuous Belief Functions for Evidential Reasoning." In Proceedings of the National Conference on Artificial Intelligence, pp. 308-313, 1984.

T1 Thompson, T. "Parallel Formulation of Evidential-Reasoning Theories." In Proceedings of the Ninth International Joint Conference on Artificial Intelligence, 1985 (submitted).

W1 Weymouth, T., et al. "Rule-Based Strategies for Image Interpretation." In Proceedings of the National Conference on Artificial Intelligence, pp. 429-432, 1983.

Y1 Yakimovsky, Y., and Feldman, J. "A Semantics-Based Decision Theory Region Analyzer." In Proceedings of the Third International Joint Conference on Artificial Intelligence, pp. 580-588, 1973.

Z1 Zadeh, L. "Outline of a New Approach to the Analysis of Complex Systems and Decision Processes." IEEE Trans. on Systems, Man, and Cybernetics, 3(1), 28-44, 1973.

Z2 ------ "A Theory of Approximate Reasoning." Machine Intelligence 9, 149-194, J. Hayes, D. Michie, and L. Kulich, eds., Wiley, New York, 1979.

END

FILMED

5-85

DTIC